M

# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

## CONTENTS

*The Relationship of Anxiety Level to Performance in Serial Learning:* J. A. TAYLOR AND K. W. SPENCE 61

*Visual Recognition Thresholds as a Function of Word Length and Word Frequency:* E. MCGINNIES, P. B. COMER, AND O. L. LACEY 65

*The Effects of Symbols, Shift, and Manipulation upon the Number of Concepts Attained:* R. S. DAVIDON 70

*Studies of Distributed Practice: VII. Learning and Retention of Serial Nonsense Lists as a Function of Intralist Similarity:* B. J. UNDERWOOD ..... 80

*Effect of Amount and Distribution of Warming-Up Activity on Retention in Motor Learning:* R. J. SILVER ..... 88

*Decrements and Recovery from Decrements in a Simple Work Task with Variation in Force Requirements at Different Stages of Practice:* E. A. BILODEAU ..... 96

*Level of Repetition in the "Spread of Effect":* F. D. SHEFFIELD AND W. O. JENKINS ..... 101

*Massing and Spacing Phenomena as Functions of Prolonged and Extended Practice:* E. A. BILODEAU 108

*A New Method for Analyzing Printed English:* E. B. NEWMAN AND L. J. GERSTMAN ..... 114

*Experiments on Sensory-Tonic Field Theory of Perception: V. Effect of Body Status on the Kinesthetic Perception of Verticality:* S. WAPNER AND H. WERNER ..... 126

*Thresholds of the Human Eye for Electric Stimulation by Different Wave Forms:* J. W. GEBHARD 132

THE JOHN CREHAR LIBRARY
RECEIVED

American Psychological Association

Vol. 44 No. 2 August 1952

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. No free reprints are supplied to authors. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

**Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49, (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *S*s, *S*'s, and *S*s'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.**

---

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 44, NO. 2 AUGUST, 1952

# THE RELATIONSHIP OF ANXIETY LEVEL TO PERFORMANCE IN SERIAL LEARNING

JANET A. TAYLOR AND KENNETH W. SPENCE
*Northwestern University* *State University of Iowa*

A number of recent studies (**3, 4, 6**) have been concerned with an attempt to study the effects of different levels of motivation on performance in a variety of learning situations. In the first of these studies Taylor (**6**) proposed the hypothesis that (*a*) the total effective drive strength (*D*) of a subject was, in part, a function of the level of internal anxiety or emotionality of the subject, and (*b*) the level of this internal emotional state would, in turn, be reflected by the responses made in a test of so-called *manifest* anxiety. This latter test consisted of items taken from the Minnesota Multiphasic Personality Inventory that had been judged by clinical psychologists to be indicative of a carefully defined behavior syndrome, manifest anxiety. The effect of different drive levels on performance in classical eyelid conditioning was then studied by comparing the conditioning curves of two groups of *Ss*, one of which had scored at the high end of the scale and the other at the low end. The results showed that the anxious (high drive) group was markedly superior to the nonanxious (low drive) group in the amount of conditioning exhibited. A subsequent study by Spence and Taylor (**5**) confirmed this finding.

On the assumption that scores on the manifest anxiety scale reflect a difference in *D*, the findings of these two studies are in line with Hull's theoretical formulation (**1**) that response strength (*R*) will be a function of excitatory potential (*E*), which is in turn a function of *D* and a hypothetical learning construct *H*. More completely the theoretical formulation may be represented as follows:

$$R = \text{f}(E)$$
$$E = \text{f}(H \times D)$$
$$D = \text{f}(\text{Anxiety score})$$
$$H = \text{f}(\text{No. conditioning trials})$$

It is obvious that the higher the anxiety score the stronger will be the excitatory potential and the greater the response strength. It should be noticed, however, that the theoretical implication of increasing drive level of *S* will depend upon the nature of the behavior situation. In situations in which there is but a single response tendency, increase in the strength of *D* results in a higher value of *E* and hence implies a stronger response. This is the state of affairs in classical conditioning, in which, by virtue of the control of the stimulus conditions, there is but a single response. In more complex learning situations such as

trial-and-error or selective learning, however, in which there are a number of alternative, competing stimulus-response tendencies, the effect of increasing drive strength will depend upon this initial response hierarchy and the relative habit strength of the correct (goal-attaining) response in the hierarchy.

Under certain conditions, particularly when the habit strength of the the correct response is weaker than those of the competing responses, raising of the drive level would be expected to have the opposite effect to that found in classical conditioning, i.e., impairment of performance. This follows from the assumption that drive ($D$) will multiply the habit strengths of all responses, correct or incorrect, thus increasing the amounts by which the excitatory strengths of the incorrect-response or responses exceed that of the correct response. The effect of this, it will be apparent, would be to lower the probability of occurrence of the correct response, thus increasing the number of errors. A larger number of trials would also be required before the stronger, incorrect responses would be weakened below the strength of the correct response.

Still another factor that needs to be considered is that raising the drive level of $S$ would also have the effect of increasing the number of responses with superthreshold excitatory tendencies, thus increasing the number of possible competing, erroneous responses. While these new superthreshold responses would probably be weaker in habit strength than the correct response, the operation of oscillation ($_{S}O_{R}$) would lead to their being stronger on some occasions. Thus, even in instances in which the habit strength of the correct response is stronger than any of the alternative superthreshold responses, the effect of increasing the drive, and hence the number of superthreshold responses, would be to increase the number of errors.

A study by Montague (3) comparing the rote serial learning of anxious and nonanxious *Ss* provides evidence bearing on these hypotheses. Montague constructed three serial lists of nonsense syllables in which the intralist similarity and association values of the syllables were manipulated in such a manner as to vary the likelihood and strength of competing responses in the different lists. On the list in which the number and strength of the competing responses should have been relatively high in comparison with the correct response, the performance of the anxious group was significantly inferior to that of the nonanxious group. The relative performance of the groups was reversed, however, on the list in which an attempt was made to eliminate or minimize the strength and number of competing associations. These results, it will be seen, are in line with the implications of the theoretical model outlined above. On the third list, which fell midway between the other two in difficulty, the two groups of *Ss* did not differ significantly.

The present investigation represents a further study of the effect on performance of differences in drive strength, defined in terms of level of scores on a test of manifest anxiety, in a learning situation that involves the presence of strong competing responses. The learning problem was a serial-learning situation involving the presentation of a sequence of choices between two responses, left and right, such as in the typical multiple T-maze situation. Instead of using one of the standard types of maze structures, however, a memory drum was employed. The $S$ was required to anticipate whether the correct response was to the left or to

the right. The appropriate response was indicated either by the word "right" appearing in the right window of the drum or the word "left" in the left window. The use of the memory drum with its 2-sec. rate of presentation of the alternatives provided for the control of the response time by forcing *S* to respond within a certain interval and thus not allowing for hesitation and possible verbal rehearsal.

## Procedure

*Subjects.*—The *Ss* were selected from an introductory psychology course on the basis of extreme scores obtained on a modified form of the scale of manifest anxiety developed by Taylor (6). All were naive with respect to the experimental task and were unaware of the reason for their selection as *Ss*.

The anxious group consisted of 20 *Ss* whose scores on the scale fell in the upper 15% of the class distribution while the nonanxious group was composed of 20 *Ss* falling in the lower 15%. The raw scores for the nonanxious group ranged from 3–11 and for the anxious *Ss* from 25–37.

*Apparatus and procedure.*—A Hull-type memory drum, mounted on a table behind a fiberboard screen, was used for presentation of the material. The learning task consisted of a series of 20 choices, typed on a continuous white tape, each choice consisting of either the word "right" or the word "left." The word "right," when it appeared, was typed on the right side of the tape, while "left" appeared always on the left. The order of the choices was as follows: LLRLRRLLRLRRRLRLLRRL.

The choices were presented serially with a 2-sec. exposure time and 2-sec. interval between exposures. The intertrial interval was 6 sec.

The instructions given to each *S* were of the usual type for serial learning. The *Ss* were required to give a response to each exposure and did so with rare exception. Trials were continued until *S* reached a criterion of two successive errorless trials.

## Results

The data on the learning in terms of errors and trials to the criterion of mastery are presented in Table 1. It will be observed with respect to both measures that the nonanxious *Ss* were superior to the anxious. Since a chi-square test indicated that the distributions of both the error and trial measures departed significantly from normality, a nonparametric test, devised by Mann and Whitney (2), was employed to test the hypothesis that there was no difference with respect to the distribution of scores for the anxious and nonanxious *Ss*. On the basis of this hypothesis the probability that the obtained differences between the two groups of *Ss* would occur was found to be between .01 and .05 in the case of the error measure and less than .01 in the case of the trial measure.

TABLE 1

Means for Error and Trial Measures

| Group | Errors | | Trials | |
|---|---|---|---|---|
| | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| Anxious | 157.55 | 17.78 | 32.75 | 2.05 |
| Nonanxious | 118.70 | 12.94 | 25.12 | 2.38 |

This finding, it will be seen, is in line with the theoretical expectation that the performance of the anxious (high drive) *Ss* would be inferior to that of the nonanxious (low drive) *Ss* in this competitive learning situation. This result is the opposite to that found in classical conditioning in which, in accord with the theory, the performance of the anxious *Ss* was superior to that of the nonanxious *Ss*.

Theoretically, the difficulty of learning the individual choice points is, in part, a function of the strength of the competing anticipatory and perseverative response tendencies. On this basis we should expect the number of errors made by anxious *Ss* to be much greater than that made by nonanxious *Ss* on the most difficult choice points, whereas this difference would be less on the points of choice intermediate in difficulty and still less and possibly even in the opposite direction in the

case of the easiest choice points. When the choice points were arranged in order of difficulty, it was observed that there was a definite tendency for the differences to be relatively greater for the difficult choice points than in the case of the easy choice points. A rank order correlation between the order of difficulty of the choice point and the corresponding magnitude of the difference between the two groups of *Ss* in terms of mean errors, was found to be .83. This value is significantly different from zero well beyond the .01 level of confidence. It was noted, moreover, that at only a single choice point, the first, did the anxious *Ss* make fewer errors than the nonanxious *Ss*. In view of the ease of learning some of the points of choice, the theory would have led us to expect a greater number of such instances. In this respect the experimental findings are not entirely in agreement with the theory. Whether the theory is incorrect or merely incomplete in the variables it specifies is not clear. Further experimentation needs to be carried out in learning situations in which the number and relative strengths of the responses competing with the correct one can be controlled and manipulated.

## Summary

The present experiment was concerned with the relation of drive strength (*D*) to performance in a learning situation. The specific theoretical implication investigated was that the performance of anxious (high drive) *Ss* would be inferior to that of nonanxious (low drive) *Ss* in a learning situation that involved the presence of competing responses, i.e., a trial-and-error type of learning situation.

Two groups of 20 *Ss* each were chosen on the basis of extreme scores made on a test of manifest anxiety, those with high scores being the anxious (high drive) group, and those with low scores being the nonanxious (low drive) group. Both groups were run in a serial learning situation involving the presentation of a series of 20 choices between two verbal responses, saying "left" or "right" at each point of choice in a memory drum setup. The criterion of learning was two successive trials in which *S* successfully anticipated the correct response at each choice point.

The results were in agreement with the theoretical expectation that the anxious (high drive) *Ss* would make a significantly greater number of errors and require a larger number of trials to reach the learning criterion.

It was also found that the points of choice that were most difficult and hence presumably involved the most competition from erroneous response tendencies provided the greatest differences between the two groups of *Ss*.

(Manuscript received December 26, 1951)

## References

1. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
2. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
3. Montague, E. K. The role of anxiety in serial rote learning. Unpublished Ph.D. dissertation, State Univer. Iowa, 1950.
4. Rosenbaum, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished Ph.D. dissertation, State Univer. Iowa, 1950.
5. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
6. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

# VISUAL-RECOGNITION THRESHOLDS AS A FUNCTION OF WORD LENGTH AND WORD FREQUENCY

ELLIOTT McGINNIES, PATRICK B. COMER, AND OLIVER L. LACEY
*University of Alabama*

Recent experiments dealing with the relationship between visual-recognition thresholds and the meaningfulness of verbal stimuli have emphasized the necessity of controlling such intraverbal parameters as word frequency. Using the Thorndike-Lorge counts as an index of relative frequency of occurrence, Howes and Solomon (1) have demonstrated that visual-recognition thresholds for tachistoscopically presented words approximate a linear, decreasing function of the logarithm of the frequency of usage. These writers indicate preference for a *response probability* variable to interferences involving perceptual mechanisms in experiments where recognition thresholds appear meaningfully related to the attitudinal significance of the verbal material employed.

Since the Thorndike-Lorge word counts are derived from a population sample and do not take into account individual frequencies of word usage, Solomon (5) has attempted experimentally to control relative word frequencies for a group of *Ss* before determining visual-duration thresholds for the words. The *Ss* leafed through cards, each presenting a nonsense word, and pronounced each word as it appeared. There were 24 nonsense words which varied in frequency of appearance from 1 to 25 times. Later, ten of these words were exposed tachistoscopically along with familiar English words and unfamiliar nonsense words, and their visual-duration thresholds were determined. Results of this investigation indicated that visual-duration threshold for words approximates a negatively accelerated decay function of frequency of word usage.[1]

It seems convincingly established that frequency of word usage is related to ease of recognition under conditions of tachistoscopic exposure. Interestingly enough, Howes and Solomon (1) in first reporting this relationship dismiss the factor of word length as having no significant bearing on their results. Although the words employed in one of their experiments varied in length from 6 to 12 letters, they made no attempt to employ this variable as an experimental parameter or to partial out its possible effect upon threshold. In fact, the investigators state, "The length of a word did not appear to affect our thresholds" (1, p. 409). In Solomon's (5) later experiment, the possible differential effect of word length was eliminated by employing only 7-letter words.

This assumption regarding the absence of a relationship between recognition thresholds and word length is rather surprising in view of the care exercised by Howes and Solomon in ascertaining the effects of such parameters as type and patterning of letters and syllable length on thresholds. Moreover, there is evidence from research by Zipf (7) that longer words occur less frequently in the language system. Although length and frequency would thus influence recognition thresholds in a similar fashion, the relative contribution of each has not been investigated.

The present experiment was planned

[1] This manuscript was very kindly made available by Dr. R. L. Solomon.

to isolate the effects of both word length and word frequency upon thresholds of recognition, as measured by duration of exposure necessary for veridical report.

## Procedure

The experiment was designed so that the results for each *S* could be summarized in a multiple-regression equation relating the three variables: word length, word frequency, and thresholds of recognition. In order to sample an adequate range of word lengths, words of 5, 7, 9, and 11 letters were selected. Each category of length contained five words selected so that their frequencies of occurrence, as listed in the Thorndike-Lorge tables (6), fell in the vicinity of 10, 100, 200, 300, and 400 per million. In determining word frequency, we followed the procedure suggested by Howes and Solomon (1) of taking the average of the figures for each word representing the Lorge magazine count and the Lorge-Thorndike semantic count. Words which satisfied the requirements of length and frequency were selected by use of Snedecor's table of random numbers (4). Twenty page numbers ranging from 1–208 (the number of pages in the Thorndike-Lorge lists) were obtained from this table. The first word on any such predetermined page which met the requirements of length and frequency was recorded, and reference was then made to the next randomly selected page. If no word of the appropriate dimensions was found on the designated page, successive pages were examined until such a word appeared. This procedure was adopted in order to avoid selecting words of any particular length or frequency having structural characteristics that would systematically influence recognition under conditions of threshold exposure.

The 20 words secured in this manner were further randomized for position in the serial order in which they were presented to the *S*s, and four "trial" words having different length and frequency characteristics were placed at the beginning of the list in order to reduce initial practice effect. The resulting set of 24 words was typed in capital letters of uniform blackness on a roll of cash register paper. This roll was mounted in a Harvard tachistoscope so that the words could be advanced singly into the viewing rectangle. The stimulus words are listed in Table 1. Brightness of the stimulus field as measured by a Weston Master II Exposure Meter was .20 ft.-candles. The pre-exposure field reflected the same illumination as the stimulus field, and *S*s remained light adapted throughout the experimental session.

TABLE 1

Experimental Words Listed in Order of Presentation to *S*s*

| Words | Frequency (Per Million) | Words | Frequency (Per Million) |
|---|---|---|---|
| Flee** | 231 | Shape | 405 |
| Controversy** | 83 | Decorum | 11 |
| Dislike** | 123 | Agent | 319 |
| Fortnight** | 28 | Forgotten | 305 |
| Substantial | 98 | Bethought | 6 |
| Stain | 90 | Possibility | 318 |
| Association | 412 | Girth | 9 |
| Household | 222 | Publisher | 95 |
| Weather | 391 | Machinery | 393 |
| Imagination | 209 | Element | 305 |
| Tightly | 106 | Organ | 196 |
| Counterpane | 7 | Inspire | 196 |

* Half of the *S*s received the words in the order given; the other half received the four Trial Words and then the experimental words in the reverse order beginning with "Inspire."
** Trial words.

The experiment was performed in a room equipped with sound-absorbing material on the walls and ceiling. Twenty *S*s, 12 men and 8 women, were drawn from students at the University of Alabama. All were naive with respect to the problem under investigation. When seated before the viewing window of the tachistoscope, *S* was instructed as follows:

> Seat yourself comfortably so that you can look through this window. I am going to show you some words at rather short exposures, and at first you may not be able to recognize them. As soon as you think you see a word, however, make a guess. The word will appear where you see the small rectangle. In order to avoid eye strain, don't stare at the rectangle all the time. I will give you a "ready" signal just before exposing the word. Be sure to guess the moment you think you recognize a word.

Wide differences in visual acuity were avoided by dismissing *S*s whose average duration thresholds for the four trial words exceeded .20 sec. Only three *S*s were so disqualified. Thresholds of recognition for the stimulus words were obtained by exposing each word first for a duration of .01 sec. and then for durations increased successively by .01 sec. increments until *S* reported the word correctly. Fatigue was reduced by allowing each *S* to rest for a few minutes after thresholds for the first 12 words had been determined. Any differential effects of practice or serial effects on the thresholds of words appearing late in the list were balanced out by reversing

the order of presentation of the 20 experimental words for ten *Ss*.

The perceptual "guesses" or hypotheses volunteered by *Ss* following prerecognition exposures of the words were recorded for later analysis.

## Results and Discussion

*Threshold relationships.*—The first step in analyzing the data consisted of computing for each *S* an equation defining the regression of threshold upon both length and frequency.[2] From these, an equation representative of the group was obtained by averaging the intercept and slope statistics for the individuals. Thus:

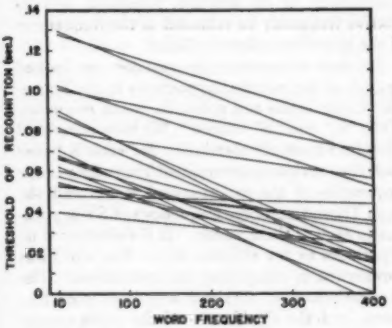


Fig. 1. Regression lines for the 20 *Ss* relating recognition thresholds to Thorndike-Lorge frequencies of the stimulus words

$Y = 5.48 - .012X_1 + .43X_2$. Here, $Y$ is the predicted threshold in hundredths of a second; the constant is the intercept value; $X_1$ denotes frequency of occurrence; and $X_2$ designates word length in letters. In each case, the 20 partial regression coefficients yielding the means were of like signs. The standard deviations of the means were as follows: mean intercept, .615; mean regression coefficient for frequency, .0012; mean regression coefficient for length, .068. All of the means differed significantly from zero at better than the 1% level of confidence.

[2] A correlation of .19 was found to exist between the length of the experimental words and their relative frequencies. This value was used as a correction constant in computing the regression equations.

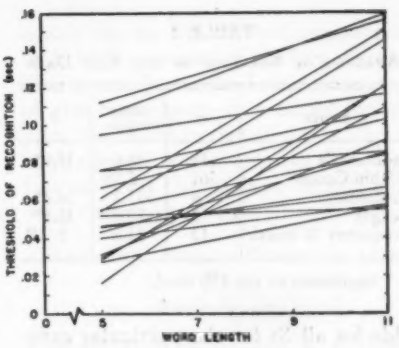


Fig. 2. Regression lines for the 20 *Ss* relating recognition thresholds to length of stimulus words

In order to demonstrate individual differences in trend as well as absolute levels of thresholds, regression lines have been plotted for the 20 *Ss* showing the predicted recognition thresholds for words of different frequencies and different lengths. These are found in Fig. 1 and 2. It should, of course, be emphasized that the obtained thresholds hold only under the specified experimental conditions.

The general results of the experiment are summarized in Table 2, showing the mean recognition thresh-

TABLE 2

Mean Recognition Thresholds in Hundredths of a Second for All *Ss*

| Frequency (Per Million) | Length in Letters | | | |
|---|---|---|---|---|
| | 5 | 7 | 9 | 11 |
| 10 | .063 | .103 | .096 | .126 |
| 100 | .049 | .055 | .056 | .066 |
| 200 | .067 | .074 | .066 | .059 |
| 300 | .047 | .059 | .062 | .056 |
| 400 | .041 | .051 | .058 | .047 |

TABLE 3

ANALYSIS OF VARIANCE OF THE RAW DATA

| Source | *df* | Mean Square | *F* |
|---|---|---|---|
| Individuals | 19 | 94.6 | 15.8* |
| Within Group | 361 | 5.98 | |
| Frequency | 4 | 325.2 | 54.4* |
| Length | 3 | 70.7 | 11.8* |
| Frequency × length | 12 | 35.8 | 5.99* |

* Significant at the 1% level.

olds for all *S*s for the particular categories employed. Inspection suggests at once a rise in thresholds with increasing length and a fall with increasing frequency. Analysis of variance of the raw data yielded results as shown in Table 3. As indicated, all effects were significant at far better than the 1% level. The significant interaction between frequency and word length may be described as follows: (*a*) an increase in frequency lowers the thresholds of long words more than those of short words; (*b*) longer words are accompanied by higher recognition thresholds to a greater extent in the case of low-frequency words than in the case of high-frequency words. In short, one cannot state the precise effect of either word frequency or word length upon thresholds of visual recognition without considering the interaction between the two.

The experimental results thus confirm the general findings of Solomon (5) and Howes and Solomon (1) with respect to the dependence of recognition thresholds upon frequency of occurrence of the stimulus words in the language system. Our results, however, not only contradict the contention of Howes and Solomon that recognition thresholds are not affected by word length, they also indicate that length and frequency interact to a significant extent. The frequency effect is most apparent with long words, while the length effect is most striking with words of low frequency.

*Hypothesis analysis.*—As noted earlier, we kept a record of the "hypotheses" volunteered by *S*s following prerecognition exposures of the stimulus words. Other investigators (2, 3) have observed that the hypotheses formed by *S*s attempting to recognize words at subthreshold-exposure durations are meaningfully related to the stimulus words. That the associational tone of the stimulus word, particularly if it has taboo connotations, influences the nature of the related hypothesis seems established. Regardless of the meaning of the stimulus word, however, perceptual hypotheses tend to be organized around the structural characteristics of the stimulus. We might ask a further question: Since the frequency of occurrence of a word is a partial determinant of its duration threshold, will its relative frequency be reflected in the frequencies of the hypotheses that it elicits?

In order to answer this question, we looked up each of the recorded hypotheses in the Thorndike-Lorge tables and noted the mean frequency of the "L" and "S" entries. We then computed a product-moment correlation for each *S* based upon the previously determined Thorndike-Lorge frequencies of the experimental words and the mean Thorndike-Lorge frequencies of *S*'s hypotheses to each of these words. If *S* volunteered no hypotheses to any stimulus word, that word was disregarded in computing the correlation. The 20 correlations so obtained were converted to *z* scores, and the significance of the mean *z* score was evaluated by means of the *t* test. The mean of the 20 *z* scores was −.121. This does not deviate significantly from zero. We may therefore conclude that there is no systematic relationship between the semantic frequency of stimulus words and the frequencies of the hypotheses elicited by these words.

This finding is probably explained by the fact that there was a significant tendency on the part of *S*s to formulate hypotheses which in general had higher Thorndike-Lorge frequency listings than those of the related stimulus words. The average frequency of occurrence of all the experimental words to which hypotheses were volunteered was 169 per million. The average frequency of the hypotheses made in response to subthreshold exposures of these words was 438 per million. The difference is significant at the 1% level of confidence. Regardless of the frequency of occurrence of a stimulus word, *S*s apparently formulate hypotheses that are relatively common in everyday usage. In addition, *S*s in the present study followed the usual tendency of basing their hypotheses primarily upon the structure of the stimulus words. This was

true whether the stimulus word was of low or high frequency.

## Summary

Thresholds of recognition for words varying in length and frequency were determined for 20 *Ss*. Despite individual differences in slope, and level of perceptual acuity, regression equations determined for each *S* indicated that duration thresholds for neutrally-toned words are a linear, decreasing function of word frequency and a linear, increasing function of word length. These findings hold true for words varying from 5 to 11 letters in length and from 10 to 400 occurrences per million in frequency. Analysis of variance of the data also revealed a significant interaction effect between frequency and length. This relationship seems to be one in which an increase in frequency lowers recognition thresholds more for long words than for short words. An increase in word length, on the other hand, raises thresholds more strikingly for low-frequency words than for high-frequency words.

No significant relationship was observed between the Thorndike-Lorge frequencies of the stimulus words to which the *Ss* volunteered hypotheses and the frequencies of the hypotheses. The mean Thorndike-Lorge frequency of the hypotheses formulated by *Ss* prior to correct recognition was significantly larger, however, than the mean frequency of the stimulus words eliciting these hypotheses.

(Manuscript received October 9, 1951)

## References

1. Howes, D. H., & Solomon, R. L. Visual duration thresholds as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
2. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
3. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abn. soc. Psychol.*, 1947, **43**, 142–154.
4. Snedecor, G. W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1946.
5. Solomon, R. L. Visual duration thresholds as a function of experimentally established frequencies of verbal stimulus-response occurrences. Paper read at the 1951 meetings of the Eastern Psychological Association.
6. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia Univer. Press, 1944.
7. Zipf, G. K. *The psycho-biology of language*. Boston: Houghton Mifflin, 1935.

# THE EFFECTS OF SYMBOLS, SHIFT, AND MANIPULATION UPON THE NUMBER OF CONCEPTS ATTAINED [1]

ROBERT S. DAVIDON

*Champlain College, State University of New York*

In recent years it has been recognized that the quality and effectiveness of an individual's thinking is related to the language or symbol system he uses as well as to his individual capacities and skills. In problem solving the particular symbols used will limit or give direction to the types of solutions that will be employed (7). Similarly, in the conceptual organization of a number of symbols representing common objects, the type of symbols used may influence the number and quality of the concepts attained. The psychologist has reason to find measures of how effectively persons can use various symbols to organize the objects of their experience. If all symbols are classified in a systematic way, it then becomes possible to establish whether there is a relationship between variations in symbols and differences in the concepts that individuals attain. One systematic way for classifying all symbols is that based upon their *level of abstraction.*

In an unpublished exploratory experiment by the writer two levels of abstraction were represented by a verbal and a pictorial representation (symbol) for each of a group of familiar objects. The names, such as "coal shovel" and "leather glove," were considered to be the more abstract, to represent wider classes of referents, while the photographs of the corresponding objects were accepted as less abstract because they offered many specific, concrete details. Significantly more concepts were attained with the pictorial symbols than with the verbal ones, and these results have led to the present development and testing of a general theory of concept attainment.

Although experimentalists have studied behavior variously referred to as concept formation, concept attainment, generalizing abstraction, and abstract and concrete behavior, there has been relatively little experimental study of the day-to-day activity whereby persons use symbols to organize and differentiate common events and objects. A general theory is required, a theory which is not confined to the acquisition of ordinary concepts during childhood or to the formation of new nonsense concepts in the laboratory.

The term concept formation implies that concepts are formed once by a special learning process, after which they are merely remembered and used over and over. This is surely an oversimplification, for although one may learn to relate a word (symbol) to a standard definition, each time he effectively uses the word the psychological event and the meaning will be somewhat different. More than memory is involved, for an organization of the whole cognitive field is necessary. The effectiveness of the individual's behavior is dependent upon the extent of his discriminatory organization, not

[1] Based upon a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, in the Department of Psychology of the University of Pennsylvania. The writer is indebted to Dr. Francis W. Irwin, under whose friendly supervision the research was conducted, and to Dr. Samuel W. Fernberger, Dr. James C. Diggory, and Dr. Mortimer Garrison, Jr., who served on the advisory committee. Miriam B. Davidon cheerfully contributed many hours to the scoring of results.

merely upon which concept he is able to recall.

If one considers the perceptual organization in first learning to use a symbol as others use it to be similar to the organizational process through which one later applies the same symbol to a specific situation, the distinction between concept formation and the use of concepts becomes of less importance. With the assumption that these psychological processes are essentially similar, one may relate the present experiment involving the organization of familiar objects to previous experiments involving the formation of new concepts.

Most of the research in concept attainment has been accomplished by two groups of psychologists whose methods and aims have been quite different. The first group has used memorization and serial learning techniques to study different stimulus variables, and has expressed results largely in terms of the number of learning trials. The experiments of Hull (**8**) and of Heidbreder (**4, 5, 6**) are representative of this group. The second group, in which Goldstein (**2**) and Hanfmann (**3**) are included, has stressed qualitative differences in response variables, such as "concrete and abstract attitudes" or the "ability to shift," and has used sorting techniques to appraise the individual's behavior.

The experiment to be described reflects the influence of both points of view. The quantitative effects of stimulus variables are emphasized, although serial learning techniques are rejected. Concept attainment is not conceived as a simple associative learning process or a form of discriminative conditioning which can be measured by the number of exposures or the time elapsed. Rather, a situation has been defined in which *S* can exercise abstract organizational ability, in which concepts are not rigidly limited by the nature of the stimuli.

## Hypotheses

I. On the average a normal adult will attain fewer different concepts when the symbols representing common objects are at a higher level of abstraction than he will when the symbols are at a lower level of abstraction.

II. (*a*) Fewer concepts will be attained if an individual has to make an initial *shift* from one conceptual organization of a group of symbols to another organization than if he begins the task without the first "irrelevant" organization. (*b*) If the previous hypothesis is verified, then the higher the level of abstraction of the symbols, the easier it will be to shift from the initial organization to the subsequent one.

III. (*a*) If one can physically separate groups of symbols and can change the stimulus field through repeated voluntary manipulations, his sorting behavior will be improved. (*b*) And if manipulation increases the number of concepts attained, then the lower the level of abstraction of the symbols, the greater will be the advantage of manipulation.

## The Independent Variables

*The level of abstraction of the symbols.*—In addition to the direct perceptual organization obtained in the discrimination of a given picture or written name there is the recognition in certain situations that both the picture and the name of an object are symbols. The symbols represent a class each member of which has certain similarities—and has differences which are ignored. In their function as symbols pictures and words need not be considered as different kinds of stimuli, but rather as symbols representing different levels of abstraction on a single continuum. The level of abstraction is determined by the number of defining common characteristics related to the symbol. As Welch has indicated, the more inclusive the class the fewer are the characteristics common to its species. "The more inclusive the class the more abstract the concept" (**10**, p. 202).

Conversely, the more detailed and more numerous are the defining characteristics the smaller is the class of referents to which the symbol pertains, and the "lower" is its level of abstraction. Theoretically, a detailed picture accepted as a symbol relates to a more limited class than does a name or short title for a corresponding object. To delimit verbally the same class symbolized by a picture might require many pages of description—and perhaps it could never be done exactly. Different persons, however, may interpret the same verbal and pictorial symbols at many different levels of abstraction. In this experiment it was assumed that for most *S*s the interpreting of the generality of the words and of the pictures followed the theoretical relationship described above.

Four levels of abstraction were used. Short titles were considered to be more inclusive than long titles, long titles more inclusive than drawings, and drawings more inclusive than photographs.

*Shift.*—Hanfmann and Kasanin (3) have presented evidence that there are differences in the ability of normal *S*s to voluntarily restructure their perceptual field. But, success in shifting may not be only a function of individual abilities but may be in part determined by the level of abstraction and the configuration of the symbols to be reorganized.

In several experiments performed previously by others (1, 3) it has been determined whether or not *S* is able to shift voluntarily, to reorganize the stimuli when asked to do so by *E*. The only quantitative measure usually taken is the number of shifts that can be made with the same stimuli. In the present experiment the effect of an initial shift was indirectly measured by comparing the performance of those who were presented with a field which was not conceptually organized, with the performance of those who had had to restructure a completely organized stimulus field.

*Manipulation.*—It was anticipated that the number of concepts that would be attained was in part a function of the spatial relationships between the objects. This followed from the assumption that the nearer two stimuli were in space (everything else being equal) the more likely it would be that they would be classified together. And, it was assumed that the fewer stimuli there were to be considered the more readily could one perceive existing relationships between them. By allowing some *S*s to manipulate the symbols, to move the cards about, smaller distances between them would be frequently attained, and by distinctly separating completed groups from the remaining field the further organization of the remaining symbols would be made easier as their number decreased.

This variable might also be described in terms of difference in isolation. For, when the positions of all symbols were fixed and there were irrelevant symbols between related ones, it was difficult to isolate the irrelevant symbols, to be constantly aware of their essential differences.

Should one consider isolation and manipulation to be separate variables, then their effects in this experiment would be confounded. But how isolation might be measured independently of grouping is not readily apparent. King's factor analysis (9) indicated that the tests in his battery were not effective for making such a distinction.

## Method

The same familiar objects were presented to every *S* by means of words or pictures, and the spatial arrangement of these was exactly the same. The *S*'s task was to find and label as many mutually exclusive three-card groups as he could within the timed period. The score was the number of concepts attained which met the criteria.

Half of the *S*s saw cards with names and half saw cards with pictures of corresponding common objects. And of those who grouped names, half grouped "long names" and half grouped a corresponding set of "short names." Also, of those who grouped pictures, half worked with photographs and half used simple line drawings.

Half of all *S*s were told initially that each of the six columns of cards represented a class of objects. These five-object classes were named by *S* when possible; if not, they were explained by *E*. The other *S*s were not informed of this "irrelevant" grouping, but they were asked at the end of the experiment whether they had observed such relationships.

Half of all *S*s were encouraged to move the cards about on the board and to put all cards of a group together in a specially provided frame before they recorded their title for the group. The other *S*s were not permitted to move the cards, but were instructed to write the title of a group and the numbers of the respective cards on the record sheet.

There were 16 conditions with eight *S*s treated alike under each condition. The *S*s were assigned to conditions at random.

## *Materials*

*Stimulus objects.*—Thirty familiar objects were selected so that it was possible to classify them into ten mutually exclusive groups of three objects each. There were four different packs of cards which represented these "same" objects. The objects were: lantern, clock, canoe paddle, egg beater, umbrella, stool, pail, eyeglasses, bi-

cycle, refrigerator, brassiere, pencil sharpener, pen, searchlight, basket, scale, curtains, safe, washtub, speedometer, saddle, cupboard, awning, typewriter, cigarette lighter, camera, fishing reel, knife, ear muffs, fan.

In the first pack there was a printed photograph representing each object, mounted on a white 3 × 5-in. card. The photographs were cut from mail-order catalogs. In the second pack there was a drawing of each object made in black ink with an A-4 Speedball pen on a white 3 × 5-in. card. The outline of the drawing was identical to that of the photograph, and a minimum of detail was included. In the third pack a "long name" (two or three words) for each object was printed on a white 3 × 5-in. card. And in the fourth pack a "short name" (one or two words) was printed in the same way. In every case the "long name" included the "short name" plus one or two adjectives, e.g., *alarm* clock. The "short names" are given in the preceding paragraph. The list of "long names" is: kerosene lantern, alarm clock, wooden canoe paddle, hand egg beater, lady's umbrella, high stool, steel pail, silver frame eyeglasses, man's bicycle, modern refrigerator, "uplift" brassiere, wall pencil sharpener, old fountain pen, motorboat searchlight, basketball basket, household scale, window curtains, floor safe, round galvanized washtub, auto speedometer, western saddle, china cupboard, striped awning, portable typewriter, table cigarette lighter, "candid" camera, casting fishing reel, butcher knife, woolly ear muffs, and electric fan.

There was also a practice set of six objects, made up similarly with four packs.

The four packs of cards were prepared to represent four levels of abstraction. For example: one of the photographs represented a particular brand and style of a round-faced spring-wound alarm clock. The ink drawing of the clock with many details of the face omitted was still readily recognizable as a round-faced alarm clock, but it would have been difficult to correctly identify the brand, for several manufacturers have made clocks which are similar enough so that the outline drawing might have represented any one of a number of models. The drawing, then, was considered to be an adequate symbol for all of the clocks of the particular model photographed, *plus* additional similar clocks of other manufacturers. The long name, "alarm clock," was a symbol which represented all of the clocks similar to the drawing, *plus* square-faced alarm clocks, pocket alarm clocks, etc. And the short name, "clock," represented all alarm clocks, *plus* grandfather clocks, office clocks, auto clocks, etc. Since each of these symbols—the photograph, the drawing, the "long name," and the "short name"—included a wider class of objects than the preceding one, each was considered to be at a higher level of abstraction.

*Sorting board.*—In order to keep the distances between the cards constant for all *S*s throughout the experiment a slotted sorting board was used. The plywood board 37 in. by 24 in. over-all had 30 shallow slots, one for each card. Each slot was 5 in. long, 2¼ in. wide, and ¼ in. deep. There were six slots spaced ¾ in. apart in each row, and there were five rows spaced 2 in. apart. The cards rested in the slots at a slight angle, which increased their visibility to *S*.

*Frames.*—Ten black wooden frames were made, each constructed of ¼-in. by ½-in. strips. The outside dimensions of each rectangular frame were 4 in. by 17½ in. The frame fitted easily over any three adjacent slots (cards) in a row.

## *Procedure*

While *S* was being given instructions, the sorting board was covered so that he could not see the 30 cards which were already distributed in the appropriate spaces on the board. Each *S* used a mimeographed record sheet and a duplicate was maintained by *E*. If *E* suspected that *S* had made an error in recording a card number, or if *E* could not read *S*'s title, he verified the entry as soon as it was made. For the last 64 *S*s, *E* recorded the time to the nearest minute when each concept was recorded.

Prior to the experiment each *S* was asked whether anyone had spoken to him about the details of the experiment. All *S*s stated that no previous information had been received. All *S*s were requested not to communicate with others regarding the experiment.

The *E* then gave the following instructions: "On this board before you are 30 cards with pictures (names) representing common objects. If you do not know what any one of the objects is, ask me about it when you first see the cards. I would like you to see how many separate groups of *three* objects you can find. When you have found three objects that belong together, write the numbers of the cards on this record sheet, here, and write your name or title of the group over here. Now this is most important! The title should *not* apply to any other object on the board other than these three. That is, no more than three objects can be described by this title. The title may have more than one word.

"Briefly, my purpose is to learn how *many* separate groups you can discover. You will have plenty of time—25 min. I will tell you when you have but 5 min. to go.

"But before you begin, let me show you a sample of six cards. Do you see a group of three here? I can group (*E* manipulates cards and uses frames for those *S*s who will do the same

during the experiment) the boy, the girl, and the chicken, No. 5, No. 3, and No. 4 together as 'creatures that walk on two legs,' or simply 'bipeds.' And, I can group the horse, dog, and deer, No. 2, No. 1, and No. 6 together as 'four-legged animals.' I would not be able to pick just three of these cards, say No. 1, No. 2, and No. 3 and label them 'living beings,' because this title would apply to more than three of the cards. If I used the name 'domestic animals' I would have one correct group of three—the chicken, the dog, and the horse, No. 4, No. 1, and No. 2. But then I could not classify the boy, the girl, and the deer together. I would have but one group instead of two. And the aim, you recall, is to get as *many* groups as possible. Do you have any questions?

"Let me review the main rules of the problem. I will put this large card with the rules there so you may refer to them later if you wish. (1) Find as many three-object groups as you can. (2) Write the card numbers and the title of each group. (3) The title of the group should not apply to *any of the other objects on the board.* (4) An object can appear in only one group."

*Manipulation Ss.*—The special instructions were: "You may move the cards about the board from place to place, just as long as each card rests in one of the openings. As soon as you have found three cards that belong together put them in a row adjacent to one another, and put one of these black frames around them—before writing the numbers and group title."

*Nonmanipulation Ss.*—The special instructions were: "You are not to touch or move the cards during the experiment. As soon as you have found three objects which form a separate group write the card numbers and the group name on the record sheet. If you decide to change a grouping, line it out and use another row on the record sheet. Now here is the board with the test objects."

*Shift Ss.*—The special instructions were: "Notice that the five objects in each column have some relationship. What are the things in this column?" If there was no reply in 10 sec. or an inadequate title was given: "These are containers for liquid." Similarly, if *S* could not name one or more of the other columns: "These all have glass faces," "These are all sports equipment," "These are all kitchen items," "These are made with cloth," "These are all office equipment." And then, "Now, your task will be to find 3-card groups so that the titles apply only to three cards, not to four, five, or more."

*Nonshift Ss.*—At the end of the period they were asked: "When you began this task did you notice any column of five cards which was a group in itself? What would you call it (them)?"

## *Subjects*

The *Ss* were 128 volunteer students enrolled in elementary psychology at the University of Pennsylvania, 65 women and 63 men. One *S* failed to take the experiment seriously and was replaced.

## *Scoring*

The measure of *S*'s performance was the number of *separate* concepts attained. Although *Ss* searched for mutually exclusive concepts, an absolute criterion of mutual exclusiveness was considered to be impractical in scoring. The criteria of a separate concept used in scoring were as follows: (*a*) The title of the group of objects had to include all three members of the group. (*b*) Every other object on *S*'s record sheet was checked to see whether it also was included within this title. If *two* other objects were included, the concept was cancelled. If *one* other object was included, the concept was acceptable, but the outside object was marked so that the following action might be taken when necessary. (*c*) Whenever a group had two or three objects which had been scored as being included within other titles, then this concept was cancelled. For example, when *S* using "long names" grouped and labeled the objects as below,

| | |
|---|---|
| *Office equipment* | portable typewriter |
| | floor safe |
| | high stool |
| *Hand-cranked objects* | wall pencil sharpener |
| | hand egg-beater |
| | casting fishing reel |
| *Sporting goods* | man's bicycle |
| | basketball basket |
| | wooden canoe paddle |

he was credited with two separate concepts rather than three. The concept "hand-cranked objects" was cancelled because a pencil sharpener is an item of office equipment and a fishing reel is a sporting goods item. (*d*) When a concept was cancelled it was eliminated entirely from the scoring of other concepts.

Judgment was still required in the determination of which objects were included under each of the titles used. To obtain a measure of the effects of variability in judgment upon the scores obtained, the two duplicate record sheets for each *S* were scored independently by two well-trained scorers. The correlation between the scores for the 128 *Ss* by the two scorers working independently was .93. There was no case in which the two scorers differed by more than one concept. The mean for one scorer was 7.29 and for the other it was 7.44, a nonsignificant differ-

ence. The interpretation of results would be the same for the data of either scorer. The results as scored by the author are given in this report.

## Results and Discussion

As the maximum possible score was ten and the minimum score was expected to be two or three, it was recognized from the start that there would be a narrow range of discrete measures. There were five *Ss* who were given scores of ten by both scorers, and approximately one-half of the *Ss* had scores of eight or over. The "ceiling" on performance may have leveled the differences between the experimental treatments. The lack of discrimination among superior *Ss* tends to add greater import to the highly significant differences obtained.

TABLE 1

The Number of Concepts Attained

| Condition | *N* | Mean | *SD* |
|---|---|---|---|
| Short names | 32 | 7.56 | 1.09 |
| Long names | 32 | 5.90 | 1.40 |
| Drawings | 32 | 7.80 | 1.13 |
| Photographs | 32 | 7.88 | .99 |
| Manipulation | 64 | 7.56 | 1.35 |
| No manipulation | 64 | 7.01 | 1.42 |
| With shift | 64 | 7.40 | 1.51 |
| No shift | 64 | 7.17 | 1.29 |

*Type of symbol.*—An analysis of variance (Table 2) indicates that one can state with high confidence that the type of symbols used to represent a number of familiar objects does influence the efficiency with which college students can organize these objects into separate classes. Performance with "long names" is significantly inferior to that with "short names" ($p < .01$), and it is also significantly less than that with drawings or photographs ($p < .01$). But there is no convincing evidence that the "short names," the line drawings, or the photographs have differential quantitative effects.

TABLE 2

Analysis of Variance

| Source | *df* | Mean Square | *F* |
|---|---|---|---|
| Symbols | 3 | 27.8 | 21.4* |
| Shift | 1 | 1.8 | |
| Manipulation | 1 | 9.6 | 7.4** |
| Symbols × shift | 3 | 0.01 | |
| Symbols × manipulation | 3 | 1.6 | |
| Shift × manipulation | 1 | 0.6 | |
| Symbols × shift × manipulation | 3 | 2.3 | |
| Residual | 112 | 1.3 | |
| Total | 127 | | |

* $P_{.01} = 3.96$. ** $P_{.01} = 6.87$.

In view of the results one might tentatively generalize that for college students the efficiency of conceptualizing is directly related to the level of abstraction of the *verbal* symbols for objects. Such a statement, based upon the direct relationship obtained, is contradictory to Hypothesis I, which proposes an inverse relationship between the level of abstraction and the number of concepts attained.

There were, however, significantly more concepts formed with pictures than with "long names," as predicted by this hypothesis. This result and the one previously discussed, therefore, cannot be accounted for in terms of a single principle of generality of meaning (level of abstraction). In order to explain the obtained results a second variable might be considered. It is suspected that differences in the *perception* of words and pictures may have an effect upon the efficiency of their organization.

*Shift.*—There is no evidence that the initial shift in this experimental situation had any quantitative effects upon the efficiency of conceptualizing. The difference in the number of concepts obtained between those who had to shift from five-card to three-card

categories and those who immediately began the organization of the smaller groups was not statistically significant.

There are three interpretations of these results which may be considered. First, one might reject Hypothesis IIa, and tentatively accept the statement that a previous organization of perceived objects has little or no quantitative effect upon a subsequent conceptual organization of these same objects. Or secondly, one might say that some other factor in this experiment counteracted the inhibitory effects of shift. A possible factor which may have facilitated the performance of *S*s who had to shift is transfer effect. The *S* finds or learns six classifications which are satisfactory to *E*, and he may note that objects are grouped together which have a common use, a common location, or a part in common, and he may apply these relationships when confronted with the main task which followed. Finally, some *S*s may have considered the preliminary five-card grouping as a trivial diversion, as they already knew that their main task would be to find three-card groups. Such an attitude would be likely to reduce the effects of the shift.

It was anticipated that some *S*s in the nonshift group might immediately recognize the five-object categories and then would find it necessary to make the same shift as those in the other experimental group. To determine whether this had occurred, *S*s were questioned at the end of the experiment, and 52 of the 64 *S*s in this group reported that they had not observed a relationship among the five cards of *any* column. No *S* reported having recognized that every column was a related group in itself. It can be said, therefore, that generally the *S*s of this experimental group did not have to make an initial shift.

After half of the results had been collected and it was recognized that the shift was not having the predicted quantitative effects, *E* decided to record the time when each concept was recorded in order to determine whether the initial shift caused a delay in the attainment of the three-object categories. Since the total time was more than adequate for most *S*s, it was considered possible that there were immediate inhibitory effects of shift, but that these effects could be overcome after a sufficient time. However, with 34 *S*s in the shift group and 30 in the nonshift group, and by tabulating the number of concepts attained in each of the five 5-min. intervals, no significant evidence of an initial delay was obtained.

*Manipulation.*—There is convincing evidence (significant at the .01 level) that the manipulation of the stimulus objects facilitated concept attainment (See Table 2). Hypothesis IIIa has been confirmed. Those *S*s who were allowed to change the spatial configuration of the cards and to isolate the completed groups were able to attain more concepts than those *S*s who were not permitted to touch the cards.

The effect of manipulation in this experiment is interpreted as a combined effect of two factors, changes in spatial arrangement and the isolation of completed groups. The more often the cards were moved about the board the more likely it was that *S* would see a relationship between adjacent cards. As the 30 cards on the large board could not be in *S*'s single span of attention, it would seem likely that the relationships between objects would be more readily discriminated when the three cards to be related were within a single span. The *S* who could not manipulate the cards was more dependent upon memory. However, *E* observed that very few *S*s

actually moved the cards about in an exploratory fashion. Many *S*s in this group did not change the position of the cards until they had attained a concept. In fact, in spite of explicit instructions to put the cards adjacent to one another first and then to write the title of the group, many *S*s wrote the title before they put the three cards together. Another common practice was to place two related objects together and then to search the board for a third member of the group. These two practices would presumably tend to reduce the advantage of manipulation.

There was behavioral evidence that the physical isolation of the completed groups did make it easier for *S*s to concentrate on the remaining cards. The *S* who could not manipulate the cards had to keep referring back to his record sheet to determine which objects had already been grouped. Several *S*s who could not move the cards complained that they kept forgetting which cards they had already grouped.

The interaction predicted between manipulation and the level of abstraction of symbols was not evident. In fact, manipulation had a somewhat greater effect with verbal symbols than it did with pictorial ones, although the interaction was not statistically significant. This possible reverse relationship might well be tested in an experimental design which would permit a wider range of scores and which would have a higher "ceiling."

*Variety of concepts used.*—The great diversity of concepts given by *S*s under each condition lends credence to the belief that the use or attainment of concepts is an active organizational process and not merely the recall of titles for fixed patterns. There were 212 different combinations of three cards used, and in many cases for the same three cards different concepts were attained by different *S*s. For example: the refrigerator, safe, and cupboard were called "things for storage" by some *S*s, "rectangular articles" by other *S*s, and "objects with doors" by still others. The number of *S*s who used each combination of objects was tallied, and in cases where clearly more than one concept was used for the same objects the repetitions of each concept were tallied separately. Only one concept was shared by more than half the *S*s. This concept was used by 68%.

The difference in the mean number of concepts attained with names and pictures cannot be accounted for by a particular concept (or concepts) attained under one condition but not under another. There was no combination of two concepts which accounted for half of the difference between "long names" and "short names," or between "long names" and the pictured symbols.

For each kind of symbol—"short names," "long names," drawings, and photographs—there were more than 100 *different* concepts used. The differences between these numbers of different concepts were not significant by the chi-square test. However, it is important to note that *S*s working with verbal symbols attained a greater number of *unique* concepts than *S*s working with pictorial symbols. A unique concept is defined as one which was attained only once during the course of the experiment. "Long names," which gave the least number of concepts per *S*, gave the greatest number of unique concepts. There were 69 unique concepts with "long names," 56 with "short names," 43 with drawings, and 43 with photographs. For the null hypothesis $\chi^2 = 8.8$, and $p < .05$. This result suggests that the more difficult is the task of conceptual organization the

more diversified and original are the concepts attained by intelligent *S*s.

There is also evidence that *S*s who were presented with different types of symbols tended to form different types of concepts. For example, more *S*s who used the "short names" attained concepts based on common *use* than did those who used drawings. This statement is derived from an analysis of the most frequently attained concepts—those attained by eight or more *S*s. The analysis was limited to these 23 most frequently attained concepts because differences in rarely attained concepts would be less likely to be reliable, and because the analysis is exploratory rather than comprehensive. The classification of concepts was accomplished by first defining a number of types and then having the two scorers classify each of the 23 most commonly attained concepts. For example: when the refrigerator, safe, and cupboard were grouped together and called "things for storage," this concept was classified as one based on common *use*. When for the same objects the title "rectangular boxes" was used, the concept was classified as one based on common *shape*. Therefore, the classification of concepts into types was not determined by the groups of objects themselves, but by the type of relationship indicated by *S* in his title. The scorers were able to classify the 23 most frequently attained concepts into five types: use, shape, parts, action, and "other" (one concept). Other types defined, such as common location and common materials, were discarded.

From the results of this classification one may tentatively accept the statement that there is a relationship between the type of symbols and the type of concepts attained.[2] The null hypothesis can be rejected at better than the .01 level of confidence. When grouping verbal symbols there appears to be a greater tendency to attain concepts based on common *use* than when grouping pictorial symbols. And with pictorial symbols, conversely, there is a greater tendency to attain concepts based upon common *parts*. Although these relationships may be a function of this particular group of objects, future *E*s should consider not only the effects of various symbols upon the efficiency of conceptualizing and thinking, but also their effects upon the nature of the concepts formed.

## Summary

The aim of the experiment was to determine the quantitative effects of the level of abstraction of symbols, of a shift, and of manipulation upon the efficiency of concept attainment. The measure employed was the number of separate concepts attained by *S*. An analysis also was made of the specific concepts attained under each experimental condition.

Four different sets of symbols which represented the "same" group of familiar objects at different *levels of abstraction* were used. These were photographs, line drawings, "long names," and "short names." The symbols in each set were presented simultaneously and no memorizing was involved. Possible proactive inhibitory effects of a *shift* in conceptual organization were investigated by requiring some *S*s first to label each column of five objects, and then to perform the main

[2] A table indicating the number of subjects of each experimental group who attained each of the twenty-three most frequently attained concepts, with concepts classified according to their type, and the definition of concepts may be obtained by ordering Document No. 3626 from the American Documentation Institute, 1719 N St. N. W., Washington 6, D. C., remitting $1.00 for 35 mm. microfilm or $1.00 for photocopies (6 by 8 in.).

task—to find as many separate groups of three objects as they could. Finally, some *Ss* were permitted to *manipulate* the cards with the symbols, i.e., to change their spatial relationships and to isolate completed groups.

With a sample of 128 college students it was found that:

1. The *Ss* attained more separate concepts with pictures of common objects than they did with the "long names" of these objects.

2. They attained a greater number of concepts with "short names" than they did with the longer, more specific names. This result contradicted the hypothesis that the lower is the level of abstraction of the symbols to be grouped, the more concepts an individual will generally attain.

3. There was no evidence that an enforced initial shift in conceptual organization had a quantitative effect upon the efficiency of conceptualizing.

4. More concepts were attained when manipulation was permitted than when *Ss* did not move the symbols.

5. The *Ss* tended to form different types of concepts with different types of symbols, and there was more diversification of concepts the more difficult the task of conceptual organization.

(Manuscript received October 27, 1951)

## References

1. Bolles, M. M. The basis of pertinence. *Arch. Psychol., N. Y.*, 1937, No. 212.
2. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **33**, No. 239.
3. Hanfmann, E., & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, Series 67.
4. Heidbreder, E. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173–189.
5. Heidbreder, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.
6. Heidbreder, E. The attainment of concepts: III. The process. *J. Psychol.*, 1947, **24**, 93–138.
7. Helson, H., & Helson, H. B. Some common features of concrete and abstract thinking. *Amer. J. Psychol.*, 1946, **59**, 468–472.
8. Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).
9. King, W. R. Ability to abstract. In F. Mettler (Ed.), *Selective ablation of the frontal cortex*. New York: Paul V. Hober, 1949. Pp. 218–238.
10. Welch, L. A behaviorist explanation of concept formation. *J. genet. Psychol.*, 1947, **71**, 201–222.

# STUDIES OF DISTRIBUTED PRACTICE: VII. LEARNING AND RETENTION OF SERIAL NONSENSE LISTS AS A FUNCTION OF INTRALIST SIMILARITY

BENTON J. UNDERWOOD[1]

*Northwestern University*[2]

It is a well-established fact that the learning of lists of serial nonsense syllables is facilitated by distributed practice (e.g., **4**, **8**, **9**). It has also been demonstrated that increasing the length of a list (**6**), increasing the rate of presentation of the items from 2 sec. to 1 sec., (**4**), or decreasing the meaningfulness (**9**), will tend to enhance the effects of distributed practice. Thus, it might seem that the more difficult the learning task the greater the probability that facilitation in learning will result from distributed practice. However, it has never been shown that this is a general law even within the restricted task of learning serial nonsense lists. That is, while increasing the length of the list or increasing the rate of presentation may increase the facilitation in learning by distributed practice, such a relationship may be produced entirely by a factor or factors intrinsic to these two methods of changing the difficulty of the task. It is quite possible that when other variables are used to change the level of difficulty, the relationship will not hold.

One variable which should be effective in changing the learning time of serial lists of constant length is intralist similarity. The first purpose of the present experiments is to vary difficulty of serial nonsense lists by manipulating intralist similarity in order to observe whether facilitation by distribution and difficulty are directly related.

The second purpose of this investigation is to examine further Gibson's (**1**) theory of verbal learning. One of Gibson's predictions is that as intralist similarity increases, rate of forgetting increases. More specifically, Gibson holds that one important factor determining rate of learning is intralist generalization. The greater the generalization the slower the learning; differentiation among items must be established as an integral part of the learning process. Furthermore, Gibson assumes that with the passage of time following learning, the generalization tendencies will recover spontaneously to produce intralist interference at recall. The greater the amount of intralist similarity (leading to generalization) the greater will be the rate of forgetting. A previous study by the present writer (**10**), using paired adjectives, has failed to confirm this expectation, and Gibson's (**2**) own experiment (also using paired associates) neither supports nor denies the theory as applied to retention. The present studies, using serial nonsense lists with different amounts of intralist similarity, will make an additional test of the Gibson hypothesis.

## PROCEDURE

*Materials.*—All lists contained 14 items, but since the first item was used only as an anticipatory cue, *S* actually learned 13 syllables. For all lists, irrespective of level of similarity, the meaningfulness value was 46.67% to 53.33% as

[1] Ross L. Morgan supervised the gathering of the data; E. James Archer is largely responsible for the statistical analyses.

[2] This work was done under Contract N7onr-45008, Project NR 154–057, between Northwestern University and The Office of Naval Research.

determined by Glaze (3). Similarity was varied in terms of frequency of repetition of letters. The complete set of rules used to construct the lists will not be presented but the method employed to vary intralist similarity must be indicated in some detail.

Each syllable has two consonants and one vowel, the first and last letters being consonants. Similarity was varied in terms of frequency of repetition of first letters and frequency of repetition of last letters. In the low-similarity lists there was no repetition of first letters and no repetition among last letters. The repetition among first and last letters was as low as possible within the limits imposed by the number of letters available. In the medium-similarity lists seven different letters were used to start a syllable, each one occurring twice. The same was true among the last letters but there was no overlap among the letters used as first and last letters. In the high-similarity lists four letters were used three times each as the first letter and one was used twice. Using different letters, the same rules held for the last letter of the syllables. Vowels remained relatively constant for all levels of similarity, each being used about equally often in each type of list.

For each level of similarity three lists were constructed. These are shown in Table 1. A given *S* learned only the lists within a given degree of similarity so that the same syllables might be used for more than one level of similarity. All lists were presented by a Hull-type drum at a 2-sec. rate.

*Design.*—Three independent experiments were performed. Each experiment dealt with a different level of similarity and differed from the others only in that respect. Within each experiment there were three experimental conditions, differentiated by the length of the intertrial rest. In one condition 2 sec. elapsed between each trial (massed learning). In the other two conditions (distributed practice) the intertrial rests were 30 and 60 sec.

With three conditions and three lists, both lists and conditions were counterbalanced by using 36 *S*s. The advantages of this design have been detailed elsewhere (11).

A common practice list was used for each group of 36 *S*s. On the practice day *S* learned this list to a criterion of seven correct responses on a single trial by massed practice, and then learned to a criterion of one perfect recitation with a 30-sec. rest between each trial. The activity used to fill this distribution interval, as well as those on the experimental days, was symbol cancellation. This task has been described previously (11).

On the experimental days the list was learned to a criterion of one perfect trial. Twenty-four hours later the list was recalled and relearned to one errorless trial. Relearning was by massed practice in all conditions. Relearning of one list and original learning of another took place at a single session.

## Results

*Practice list.*—Evaluation of practice-day performance will show the equality of the three groups of *S*s. Table 2 gives the mean number of trials required to learn the practice list to one perfect trial, along with the mean number of errors per trial. Analysis of variance has shown that the within-groups variance is larger than the between-groups variance for both trials to learn and errors per trial. It

TABLE 1

Experimental Lists

| Low Similarity | | | Medium Similarity | | | High Similarity | | |
|---|---|---|---|---|---|---|---|---|
| VOD | JIR | QYT | YUG | MIC | HEF | KIV | GEZ | DUF |
| HAX | LYB | SOZ | NOL | SUJ | KAM | FUW | WIQ | NOK |
| NIW | KON | FAH | JEP | XOW | TIS | XIR | TEZ | VAY |
| JAL | FEG | NUY | RUH | ZET | POB | JUW | BIH | HEF |
| MIQ | VAP | LER | YAX | VOK | WUS | YUL | WEP | DAK |
| FEC | BIH | VAQ | ZEP | XAN | KOR | FEC | TAJ | QIX |
| BAJ | WUS | DYP | FAH | GIJ | WEM | KAW | MIH | VEB |
| LOZ | HEX | JIS | COQ | VUW | DAF | YOR | BAJ | DUY |
| KEM | CAZ | HUJ | JEV | SOZ | LUB | JAL | MEP | VOK |
| RYG | SIQ | TOF | ZUL | BEK | POY | FOV | GIJ | HEB |
| XUN | MYD | KIV | RIQ | VUT | HAX | JIR | TOQ | NUY |
| TOQ | RUK | GAC | CEG | QOZ | DUY | XEL | WIH | HAX |
| WEP | NAS | YOB | FAX | MUN | TOF | KOC | MAQ | NEF |
| DUF | POV | XEL | NUV | GAC | LER | JEV | BUP | QIB |

TABLE 2

PERFORMANCE OF THE GROUPS ON THE COMMON PRACTICE LIST

| Group | Trials to Learn | | Errors Per Trial | |
|---|---|---|---|---|
| | *M* | σ*M* | *M* | σ*M* |
| Low-Similarity | 29.69 | 1.38 | 1.51 | .14 |
| Medium-Similarity | 28.75 | 1.83 | 1.55 | .13 |
| High-Similarity | 29.75 | 1.96 | 1.47 | .12 |

may be concluded, therefore, that the groups were initially comparable.

*Trials to learn.*—The basic results in terms of the mean number of trials required to learn to a criterion of one perfect trial with the three levels of similarity and the three intertrial rest intervals are shown in Fig. 1.

Three facts, evident from these curves, are confirmed by the statistical analysis: (*a*) As intralist similarity increases, the mean number of trials to learn increases. (*b*) Learning is faster under the two distributed conditions than under the massed condition in all experiments. The *F* is 6.16, and for 2 and 200 *df* an *F* of 4.71 is needed for the 1% level of confidence.

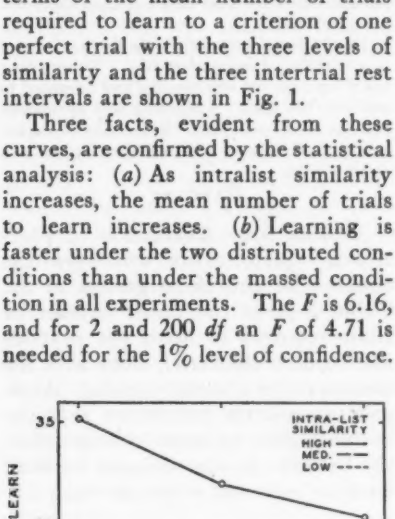


FIG. 1. Learning of serial nonsense lists as a function of intralist similarity and intertrial rest.

(*c*) The third fact, not clearly apparent in Fig. 1, is that there is little interaction between intertrial rest and intralist similarity (*F* is 1.06). Thus, while an increase in intralist similarity greatly increased the difficulty of learning, the learning of the more difficult lists is no more influenced by distributed practice than the less difficult.[3]

*Errors during learning.*—It was expected that as intralist similarity increased, overt errors per trial would increase. However, the statistical analysis gives no basis for concluding that either intralist similarity or intertrial rest is a significant source of variance as far as errors are concerned. Thus, wide differences in learning appear as a function of intralist similarity but no differences in error frequency. Actually, the data show that a somewhat greater number of errors was made in learning the medium-similarly lists than in learning the low, but the medium- and high-similarity lists produced about the same error frequency. For all conditions of intertrial rest combined, the mean number of errors per trial made in learning the low-similarity lists was 1.21. For medium-similarity the value was 1.45, and for high, 1.49.

*Recall.*—The complete statistical analysis for the recall scores is shown in Table 3. In this analysis the appropriate error term for testing similarity is *Ss*/Similarity. For testing the other main effects the proper term is Pooled *Ss* × Practice/Similarity. For evaluating list differences within a

[3] The complete statistical analysis of the learning scores will not be presented. The method used is shown in detail in Table 3 where the recall scores are analyzed. One additional finding of the present analysis may be noted. The interaction between learning-how-to-learn and similarity was significant at about the 5% level of confidence, with the practice effect being greatest for the high-similarity lists.

TABLE 3

ANALYSIS OF VARIANCE OF RECALL SCORES

| Source of Variation | *df* | Mean Square | *F** |
|---|---|---|---|
| Independent observations | | | |
| Similarity | 2 | 15.9383 | 1.280 |
| *Ss*/Similarity | 105 | 12.4526 | |
| Total between *Ss* | 107 | 12.5177 | |
| Correlated observations | | | |
| Practice | 2 | 14.7809 | 3.065 |
| Practice × Similarity | 4 | 12.4059 | 2.573 |
| Intertrial rest | 2 | 33.2994 | 6.905 |
| Intertrial rest × Similarity | 4 | 8.0494 | 1.669 |
| Lists/Similarity (pooled) | 6 | 3.6420 | |
| $\text{Lists/Similarity}_{\text{Low}}$ | 2 | 0.7778 | — |
| $\text{Lists/Similarity}_{\text{Med.}}$ | 2 | 8.0371 | 2.008 |
| $\text{Lists/Similarity}_{\text{High}}$ | 2 | 2.1112 | — |
| Pooled *Ss* × Practice/Similarity | 198 | 4.8224 | |
| $Ss \times \text{Practice/Similarity}_{\text{Low}}$ | 66 | 5.6970 | |
| $Ss \times \text{Practice/Similarity}_{\text{Med.}}$ | 66 | 4.0025 | |
| $Ss \times \text{Practice/Similarity}_{\text{High}}$ | 66 | 4.7677 | |
| Total within *Ss* | 216 | 5.3457 | |
| Total for experiment | 323 | 7.7216 | |

* With 2 and 200(198) *df*, *F* at .05 level is 3.04, and at .01, 4.71. With 4 and 200(198) *df*, *F* at .05 is 2.41, and at .01, 3.41. With 2 and 66(65) *df*, the .05 level is 3.14, and the .01 level is 4.95.

given level of similarity, the appropriate term is *Ss* × Practice/Similarity for that particular level. The two most pertinent facts in Table 3 are that intralist similarity is not a significant source of variance at recall while intertrial rest during original learning is. Furthermore, with regard to the influence of intertrial rest, the data show that recall is better following massed than following distributed practice. For all experiments combined, the mean number of items recalled 24 hr. following massed learning was 5.00. For the 30-sec. distributed condition the corresponding value was 4.19, and for the 60-sec. condition, 3.94. Clearly, then, in these experiments, massed learning is followed by better recall than is distributed learning.

Several other analyses of the recall scores have been made to determine exactly on which items massed practice was superior. The first analysis concerns the relationship between number of correct anticipations (reinforcements) during learning and the per cent correct at recall. We know from the facts of the bowed serial position curve that there are wide differences in the frequency with which each item is anticipated correctly during learning. It is also possible that the probability of an item being correct at recall is related to its position in the list independent of the number of reinforcements. Therefore, if we are to compare per cent correct at recall for different intertrial rest conditions as a function of number of reinforcements, we should first ascertain that the shapes of the bowed acquisition curves during learning are comparable for the different intertrial rest conditions. This has been done and it can be categorically stated that the shapes of the three curves (when number of correct anticipations is plotted against serial position) are highly similar. In plotting these curves differences in lists due to similarity were equally represented in all three intertrial rest conditions, i.e., the three experiments were combined. It is therefore reasonable to proceed with an item analysis of recall as a function of intertrial rest during learning with confidence that there is little if any position bias confounding the analysis.

In making the item analysis for recall, the total number of correct anticipations for each item during learning was tabulated along with whether or not the item was correct at recall. For each of the intertrial rest conditions there was 1404 items (13 items per list × 36 *Ss* × 3 conditions of similarity). These have been categorized as items having 1–7 reinforcements, 8–11, 12–15, 16–20, and 21-and-over reinforcements. The number of items in each category is in no

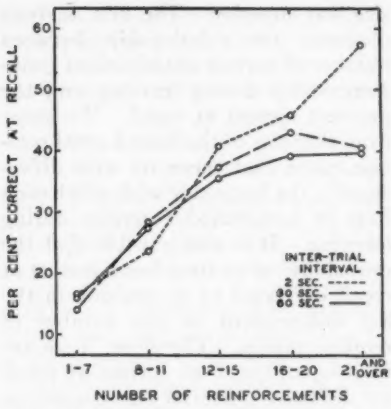


FIG. 2. Percentage correct at recall as a function of intertrial rest and number of reinforcements during learning

case less than 223. As an illustration of the computation: there were 251 items in the massed condition (2-sec. rest) having 1–7 reinforcements. Of these 251 items, 42 (16.7%) were anticipated correctly on the first recall trial.

Making the above calculations for each category of reinforcement for each condition of intertrial rest allows the plotting of the per cent correct at recall as a function of the two variables. Such a plot is shown in Fig. 2. The one clear fact demonstrated by these curves is that recall under the massed condition is largely superior only in the category where the number of reinforcements during learning is greatest (21 and over). It will be noted that the recall of items in this category is over 55% for the massed conditions and about 40% for the distributed. It might become apparent at once that this difference could be an artifact of the categorization procedure. That is, since learning under massed practice took longer than learning under distributed practice, more items having 21-and-over reinforcements would be counted for the massed conditions than for the distributed. Furthermore, the mean number of reinforcements in this category should be greater for massing than for distribution. If there is a direct relationship between number of reinforcements and per cent correct at recall, it would be expected that the massed condition would result in better recall for items in this category. Further analysis shows, however, that this is not the case. We have evaluated only those items in the 21-and-over category by groupings of two's, i.e., 21–22, 23–24, 25–26, etc. For each of these small groupings the per cent correct at recall was calculated, and in each case recall following massing was equal to or superior to recall following distribution. It must be concluded that the better recall following massed practice for highly reinforced items is not an artifact of the grouping procedure but represents a true phenomenon.

We have shown above that the difference in recall following massing and distribution is largely limited to items which have been heavily reinforced. By the very nature of the bowed acquisition curve such items must be near the first or near the last part of the list. In Fig. 3 we have plotted the

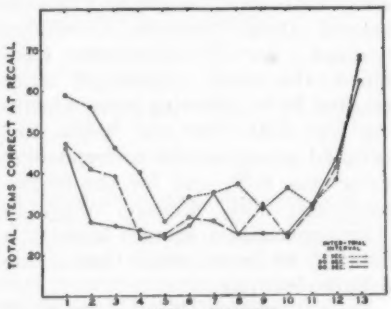


FIG. 3. Percentage correct at recall as related to serial position and intertrial rest

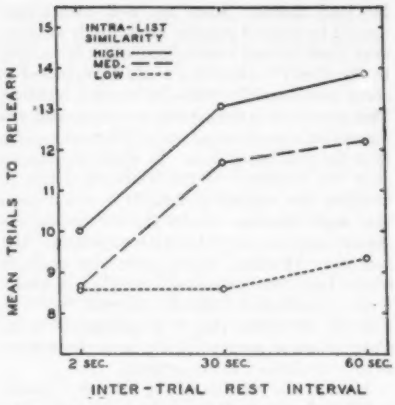


FIG. 4. Relearning as a function of intralist similarity and intertrial rest

absolute number of items recalled for each serial position. It is apparent that the massed condition obtains superiority in recall over the distributed conditions largely for items in the early portion of the list.

*Relearning.*—The mean number of trials required to relearn to one perfect trial under all conditions is shown in Fig. 4. There are three facts of note exhibited in Fig. 4: (*a*) The greater the similarity the greater the number of trials required to relearn. The *F* (6.83) is significant well beyond the 1% level of confidence. (*b*) Relearning is faster following massed than following distributed practice (*F* is 9.48). (*c*) While the curves suggest some interaction between similarity and intertrial rest, the *F* is not significant (1.61).

## DISCUSSION

Little need be said about the learning of the lists. As expected (from previous investigations) distributed practice facilitated the learning of serial nonsense lists. It may be noted, however, that no evidence was presented which showed that 60 sec. between trials was consistently superior to 30 sec. between trials; in two of the experiments the longer interval facilitated more than the shorter, in the other less. There was no interaction between intralist similarity and intertrial rest. This means that difficulty of learning *may* be increased without a corresponding increase in the facilitation produced by distributed practice. The differences in trials to learn for the different levels of similarity were clear-cut; approximately 35% more trials were required to learn the high-similarity lists than the low. Thus, while it is possible that with greater extremes in difficulty (produced by similarity) distribution would have a differential effect on learning, there is nothing in our data to suggest it. When differences in difficulty produced by the manipulation of variables other than intratask similarity have been shown to produce interaction (**4, 6, 9**), the causal agents may be specific to the variable. The other alternative is that interaction between difficulty and facilitation by distributed practice can only be demonstrated with distribution intervals longer than those employed here.

Trials to learn increased directly as a function of intralist similarity; errors per trial in learning did not. While the medium- and high-similarity lists produced more errors than did the low-similarity lists, evaluation of the error data as a whole gave us no basis on which to state a definitive relationship. It is our belief that there is a direct relationship between similarity and error *tendencies* but that *S* will only "tolerate" so many overt errors. When a maximum is reached, further increase in errors may not be expected.

The findings have demonstrated that recall is better following massed practice than following distribution. The previous experiment most comparable to the present study is that of Hovland (**5**). Using serial nonsense lists and a 2-*min.* distribution interval, he found recall to be better following distributed practice. Hovland's finding was one of the major empirical contradictions to the theory developed by Hovland, Hull, and others (**7**). The portion of the theory which deals with the present problem postulates a form of inhibition which generates during learning and dissipates with rest. (In its essential properties this inhibition is very comparable to that called *reactive inhibition.*) The effective excitatory strength between a stimulus and a response at any given moment is the algebraic sum of the positive tendency (learning) and the inhibition. According to the theory, during massed learning there would be little opportunity for the inhibition to dissipate; hence, the learning component would be considerably greater than the performance would indicate because it was being counteracted by the inhibitory component. In distributed practice, on the other hand, inhibition would remain fairly well dissipated during acquisition. Such a theory, of course, gives a fairly straightforward accounting of the fact that learning is faster under spaced than under massed practice.

However, the theory must also predict that retention following massed practice will be better than following distributed practice. This follows from the fact that given a rest interval the inhibition accruing during massed learning will have dissipated, thus leaving the greater learning component developed under massed practice to demonstrate itself in better recall performance. Hovland, as mentioned earlier, found the opposite result. This contradiction to the theory was never satisfactorily resolved.

The present results would seem (on the surface) to confirm the Hull-Hovland theory. However, automatic confirmation of their theory by the present results does not necessarily follow. We have shown that the better recall under massed, as compared with distributed, practice occurs with items having many reinforcements. Yet, it is not clear in our data whether this difference is produced because the recall of the massed items is facilitated or because the recall of the distributed items is inhibited. By the Hull-Hovland theory, the massed items should be facilitated. However, Fig. 2 more strongly suggests that the distributed items which have been heavily reinforced are inhibited. For, we may commonly expect some direct (though not necessarily linear) relationship between number of reinforcements during learning and per cent correct at recall (10). Figure 2 shows that under distributed conditions items with 21-and-over reinforcements are recalled no better than those with 16–20 reinforcements. Thus, it appears that the items which are most heavily reinforced are being inhibited in some fashion following distributed learning. Nevertheless, since we do now know the "true" relationship between number of reinforcements and recall, there is no way by which this issue can be satisfactorily settled from analyses of the present data.

Another approach to the problem is to consider the first recall trial to be a relatively unstable index of learning. We have noted that the recall difference following massed and distributed learning occurs primarily during the first part of the recall trial. This fact may indicate the operation of a very temporary inhibiting factor following distributed learning. During distributed learning, symbol cancellation came between each trial; hence, recall on each trial was always preceded by this activity. Before recall, no symbol cancellation was given for any condition. This change in context between learning and recall for the distributed conditions may have been sufficient to disrupt temporarily the recall performance. Such a possibility, of course, is subject to experimental test.

While the interpretation of the recall results must remain somewhat inconclusive, the relearning results are clearly in accord with the Hull-Hovland theory. Since the S–R associations formed by massed practice are actually stronger than those formed under distribution (according to the theory), relearning would be expected to occur more rapidly following massed learning. This was found in the present results; indeed, the massed list was relearned some 25% more rapidly than the distributed lists. In short, one cannot deny the usefulness of the inhibition theory in handling the relearning data; it is unfortunate that such clear-cut results (for or against the theory) were not found for the recall scores. The fact that Hovland found quite the opposite results from those presented here when he used a 2-min. distribution interval indicates, however, that the inhibition theory is inadequate in its present form to account for the facts of retention as they are known at the present.

Finally, the relation of the retention results to Gibson's theory (1) may be indicated. The recall results do not support the Gibson hypothesis; recall did not vary as a function of similarity. This squares with previous findings using paired adjectives (10). Gibson's theory states that generalization tendencies recover with the passage of time following learning to produce once again the interference which resulted in differences in original learning. The greater the original generalization (similarity), the greater the interference and the poorer the recall. However, the relearning scores do support the hypothesis; a greater number of trials was required to relearn the high-similarity lists than the low. And, although we did not present the data, error analyses show more errors per trial were made in relearning the medium- and high-similarity lists than in relearning the lists of low similarity. As during original learning, the medium- and high-similarity lists could not be differentiated in terms of errors during relearning. Thus, the relearning data support the Gibson theory; the recall results do not. We are unable to tell from our present results whether this latter failure is due to inadequacies in the theory or to inadequacies of the recall scores as measures of retention.

## Summary

Three experiments were run to determine: (*a*) the adequacy of the Gibson theory of verbal learning for the acquisition and retention of serial lists of nonsense syllables of varying degrees of intralist similarity, and to determine (*b*) if increasing difficulty in learning produced by increasing intralist similarity results in a correspond-

ingly greater facilitation in learning by spaced practice. Each experiment differed only in terms of intralist similarity of the nonsense syllables. Within each experiment three intertrial rest intervals were used, namely, 2, 30, and 60 sec. A separate group of 36 *Ss* served in each experiment with the three intertrial rest conditions and the lists being counterbalanced within each experiment. Under all conditions the list was learned to one perfect trial, and following a 24-hr. rest was relearned to one perfect trial by massed practice.

The results show:

1. Significant differences in rate of learning as a function of intralist similarity, with no consistent relationship between errors per trial and intralist similarity.

2. No interaction between intralist similarity and intertrial rest. Distributed practice resulted in faster learning than did massed, but this facilitation was no greater for the high-similarity lists than it was for the low.

3. Recall did not vary as a function of intralist similarity but did vary as a function of intertrial rest during learning, being greater following massed than following distributed practice. The superiority of recall following massed learning was found to be confined largely to heavily reinforced items in the early part of the list.

4. Relearning was much more rapid following massed practice than following distributed practice, and was more rapid for the low-similarity lists than for the high.

The recall results do not support the Gibson theory; the learning and relearning data do. The retention results (both recall and relearning) are in direct contradiction to previous findings of Hovland using a 2-min. distribution interval but are in fair accord with his inhibition theory. While no completely satisfactory explanation of the recall results was presented, they do not negate the Hovland theory and the relearning results definitely support the hypothesis.

(Manuscript received October 29, 1951)

## References


1. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
2. Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.* 1942, **30**, 185–200.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
4. Hovland, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
5. Hovland, C. I. Experimental studies in rote-learning theory: VI. Comparison of retention following learning to the same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, **26**, 568–587.
6. Hovland, C. I. Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27**, 271–284.
7. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
8. Patten, E. F. The influence of distribution of repetitions on certain rote learning phenomenon. *J. Psychol.*, 1938, **5**, 359–374.
9. Tsao, J. C. Studies in spaced and massed learning: III. Paired-associate and serial learning. *Quart. J. exp. Psychol.*, 1950, **2**, 13–18.
10. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
11. Underwood, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, **43**, 329–340.

# EFFECT OF AMOUNT AND DISTRIBUTION OF WARMING-UP ACTIVITY ON RETENTION IN MOTOR LEARNING[1]

REUBEN J. SILVER

*Mental Hygiene Clinic, Fergus Falls, Minn.*

It has been found by Ammons (1) and several other investigators that initial learning in certain activities is impeded by the lack of proper set. Learning is then facilitated by the acquisition of this set during the first few practice trials. Similarly, when practices ceases for a period of time, e.g., when a rest interval is introduced, there occurs a loss of the perceptual-motor adjustments which had been acquired during the performance of the task. This set must then be reinstated when practice is resumed.

The decrement in learning which is brought about by the lack of proper set and which is overcome during the initial learning or relearning trials is sometimes referred to as warm-up decrement. While investigating the characteristics of the motor learning curve following the introduction of a rest interval, Ammons hypothesized that warm-up decrement should be a negatively accelerated function, both of the amount of prerest practice and the length of the interpolated rest. Using the pursuit rotor he (2) was able to substantiate these two predictions, and further, found that warm-up decrement reached its maximum after about 8 min. of work and after about 10 min. of rest.

A different approach to the problem has been initiated by Irion (4). He reasoned that filling the rest interval with a set maintaining activity should prevent *S* from losing the perceptual-motor adjustments acquired during the performance of the task, and consequently, retention should be improved. One of his experiments (5) suggests that this is true. Using nonsense syllables as the learning task and 48 hr. as the rest interval, he had his *S*s engage in 0, 1, 5, and 10 trials of warming-up activity. The warming-up activity was introduced immediately before the recall period, and consisted of the recitation of three-place digits which were presented in a manner similar to that in which the nonsense syllables had been presented. The results of this experiment showed that retention increased up to the point of five warming-up trials but decreased when ten warming-up trials were used.

To explain this finding, it is advantageous to employ the concept of reactive inhibition ($I_R$). Hull postulates (3) that once a response is made, there is built up in the organism a tendency to avoid its repetition. This negative drive is called reactive inhibition and is a positive function of the amount of work involved in the response in question. Further, since reactive inhibition is a drive whose goal is cessation of response, it dissipates during a rest interval. In this way, one may explain the advantage of introducing a rest interval during learning, or more generally, the superiority of spaced practice.

Hull further indicates that $I_R$ probably is an effector phenomenon and does not involve the entire sensorimotor arc. Therefore, if the learning task and the warming-up activity are similar from the point of view of the effectors involved, the $I_R$ produced by both should be similar (in terms of the effectors involved), and thus, as provided by Hull's theory, it is probable that

[1] This paper is based on part of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at New York University. The author wishes to express his gratitude to Professor L. W. Crafts for his valuable aid and advice throughout the progress of the study. Thanks are also due to Dr. Hilton N. Wasserman for his valuable suggestions.

these two negative drives will summate with each other. That is, if approximately similar effectors are involved, it is justifiable to assume that the $I_R$ produced by the learning task will summate with the $I_R$ produced by the warming-up activity.

Bearing this in mind we may go a step further: Immediately following the cessation of practice, the $I_R$ accumulated during the practice period should begin to dissipate. With rest, however, another tendency results, namely, the loss of set. In many instances the beneficial effect of the dissipation of $I_R$ is more than enough to outweigh the decrement resulting from the loss of set. Consequently, reminiscence often is observed.

With the introduction of the warming-up activity, two opposing effects again result: reinstatement of set, and the accumulation of $I_R$ for the warming-up activity. If the results of Kimble (7) concerning the rapidity with which $I_R$ accumulates are correct, the effects resulting from the reinstatement of set may be unobserved because of the additional $I_R$ developed while engaging in the warming-up activity. Thus an apparent paradox is explainable as follows: Warming-up activity aids retention as long as the benefits accruing from the reinstatement of set are greater than the detrimental effects of the accumulation of additional $I_R$. After a certain point, all the perceptual-motor adjustments lost during rest are regained, and therefore, further warming-up activity can result only in the accumulation of additional $I_R$. Since the $I_R$ produced by the warming-up activity can summate with the $I_R$ resulting from the performance of the learning task, the result should be a decrease in performance.

The results of part of an experiment by Kientzle (6) tend to support this hypothesis. Using the alphabet printing task, she had one group of *S*s practice 60 sec. and rest 60 sec. The entire interval was spent in making tallies and crossbars. It would be expected that with a 60-sec. work period the optimal amount of warming-up activity would be quite small. That is, it would be expected that 60 sec. of warming-up activity would be greater than the optimal amount. This inference is supported by Kientzle's results, which indicate that retention was impaired when compared with a similar group who did nothing during the rest interval.

The significance of this discussion for the present experiment may be summarized as follows:

First, there should be an optimal relationship between the amount of prerest practice, the amount of warming-up activity, and retention. It has been indicated that when a rest interval has been introduced, *S* loses some of the perceptual-motor adjustments acquired during the performance of the task. The greater the amount of prerest practice, the greater is the number of adjustments acquired, and susceptible to loss during rest. Therefore, the amount of warming-up activity necessary to reinstate set should be an increasing function of the amount of prerest practice.

Secondly, although warming-up activity reinstates set, it also results in the accumulation of additional $I_R$. Preventing *S* from accumulating $I_R$ during the warming-up activity should result in superior retention, and this should hold true even if the entire interval is filled with the warming-up activity. The prevention of the accumulation of $I_R$ can easily be accomplished by distributing the warming-up activity. On the other hand, if the warming-up activity is massed, this should produce a decrement in performance of the learning task when practice is resumed. This assumes that the $I_R$ produced by the warming-up activity and that produced by the learning task summate, provided both activities are similar.

The above discussion leads to three predictions which the present experiment was designed to test: (*a*) The amount of warming-up activity necessary to reinstate set should be an increasing function of the amount of prerest practice. (*b*) With increasing amounts of massed warming-up activity, retention should increase up to a point and then decrease. (*c*) With increasing amounts of distributed warming-up activity, retention should increase up to a point and then level off.

## Method

*Subjects.*—The *S*s were 24 groups of 20 students from classes in general psychology at New York University, or City College of New York. The experiment was administered to one group at a time. Since the data are analyzed by means of the analysis-of-variance technique, it is essential that the number of *S*s in each group be equal (10). Although the experiment was administered to classes of unequal size, the groups were equalized by random selection.

*Materials.*—The learning task was the inverted alphabet task described by Ruch and Warren (8) and discussed by Kimble (7). This task requires *S* to print the letters of the alphabet upside down and from right to left in such a manner that when the paper is rotated 180° the letters appear properly written.

*Procedure.*—In Part I of the experiment, *S*s were divided into three groups, each having a different amount of prerest practice. The *S*s in Group 1 received a 10-min. rest after Trial 5; those in Group 2 after Trial 15; and those in Group 3 after Trial 30.

The *S*s in each group were divided into six subgroups receiving 0, ½, 1, 1½, 2, and 2½ min. of warming-up activity during the rest interval and immediately prior to the resumption of practice. The warming-up activity required *S* to make tallies and crossbars on a worksheet similar to that used for the alphabet printing task.

The group receiving 15 trials of prerest practice included three additional subgroups, which had 3, 4, and 5 min. of warming-up activity, respectively. This procedure was utilized to facilitate the comparison of the results with those of Part II of the experiment.

The classes constituting the groups having various amounts of warming-up activity in each condition were assigned at random. This procedure was designed to help assure the equivalence of the groups.

Each *S* was given two booklets, one for the experiment proper and the other for the warming-up activity. The one for the experiment proper consisted of four work sheets and a title page. Each work sheet was set up in such form that *S* was required to print each letter in a ¼-in. square box. Each sheet was divided into nine trials. The booklet for the warming-up activity was similar to the one used in the experiment proper.

Each of the experimental groups received 36 30-sec. trials and a single rest of 10 min. Following Kimble (7), *S*s were told after 29 sec. to proceed to the next trial. This same procedure was employed even on Trials 9, 18, and 27, after which the page had to be turned. The score was the number of letters printed, accuracy being disregarded.

Part II of the experiment was essentially similar to the first. Amount of prerest practice, however, was held constant at 15 trials. Three groups of *S*s engaged in 3, 4, and 5 min. of *distributed* practice on the warming-up activity during a 10-min. rest, and just prior to the resumption of practice. The *S*s alternately worked 30 sec. and rested 30 sec. during the warming-up period. The group that had 5 min. of warming-up activity began this practice at the start of the rest period. The other groups began the warming-up activity after 2 and 4 min. of the interval had elapsed.

## Results and Discussion

*Warming-up activity and prerest practice.*—All groups were equivalent on comparable prerest trials.[2] The mean increase in score on the first postrest trial for the 18 groups having various amounts of warming-up activity is shown in Fig. 1. It is apparent that after rest the performance of the *S*s in the different groups increased up to a point, as the amount of warming-up activity increased, and then declined. The point after which performance began to decline is a function of the amount of prerest practice.

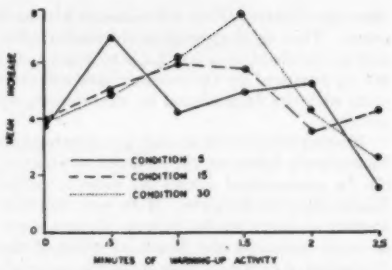


Fig. 1. The mean increase in score on the trial following various amounts of warming-up activity after 5, 15, and 30 trials of prerest practice.

In order to determine whether the trends depicted in Fig. 1 are significant, an analysis of variances was performed (Table 1). The *F* test shows the variance due to warming-up activity to be significant at the .01 level.

To evaluate these results further, the significance of the differences between the increase in score for the groups having rest and warming-up activity and those having rest without warming-up activity was computed. The *t* tests are summarized in Table 2. With five trials of prerest practice, ½ min. of warming-up activity yields the maximum retention score. After 30 trials of prerest practice, performance is significantly increased following 1½ min. of warming-up activity. Although none of the differences in

[2] An analysis of these prerest scores is presented in (9).

TABLE 1

ANALYSIS OF VARIANCE OF THE MEAN INCREASE IN SCORE ON THE TRIAL FOLLOWING VARIOUS AMOUNTS OF WARMING-UP ACTIVITY AFTER 5, 15, AND 30 TRIALS OF PREREST PRACTICE

| Source of Variance | Sum of Squares | *df* | *F* |
|---|---|---|---|
| Practice | 13.85 | 2 | 1.00 |
| Warming-up activity | 469.23 | 5 | 5.49* |
| Interaction | 262.32 | 10 | 1.54 |
| Error | 5842 | 342 | |
| Total | 6587.40 | 359 | |

* Significant at the .01 level of confidence.

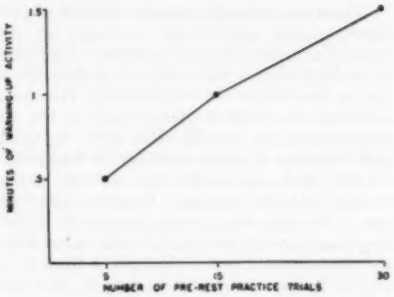


FIG. 2. The amount of warming-up activity which produced the maximum relearning score after 5, 15, and 30 prerest trials.

Cond. 15 is significant, the trend is similar and suggests that the maximum retention is found following 1 and 1½ min. of warming-up activity.

The relationship between warming-up activity and the amount of prerest practice is indicated more clearly in Fig. 2. Each point on the curve shows the amount of warming-up activity following 5, 15, and 30 trials of prerest practice which yields the greatest increase in score on the first postrest trial. It is apparent that the amount of warming-up activity which produces the greatest increase in score is an increasing function of the amount of prerest practice.

TABLE 2

MEAN DIFFERENCE IN INCREASE IN SCORE AFTER REST BETWEEN THE GROUPS HAVING WARMING-UP ACTIVITY AND THOSE HAVING NO WARMING-UP ACTIVITY

| Amount of Prerest Practice | Amount of Warming-up Activity (Min.) | Mean | *SD* | Mean Diff. | *t** |
|---|---|---|---|---|---|
| 5 Trials | 0 | 3.70 | 2.26 | | |
| | ½ | 6.90 | 4.52 | 3.20 | 2.52** |
| | 1 | 4.30 | 2.59 | .50 | <1.00 |
| | 1½ | 5.05 | 5.01 | 1.06 | 1.06 |
| | 2 | 5.30 | 3.02 | 1.26 | 1.26 |
| | 2½ | 3.70 | 4.36 | −1.65 | 1.65 |
| 15 Trials | 0 | 4.00 | 3.49 | | |
| | ½ | 5.00 | 3.15 | 1.00 | <1.00 |
| | 1 | 5.95 | 4.40 | 1.95 | 1.54 |
| | 1½ | 5.80 | 4.18 | 1.80 | 1.42 |
| | 2 | 3.60 | 3.83 | −.40 | <1.00 |
| | 2½ | 4.30 | 5.78 | .30 | <1.00 |
| 30 Trials | 0 | 3.80 | 4.43 | | |
| | ½ | 4.85 | 4.39 | 1.05 | <1.00 |
| | 1 | 6.20 | 3.46 | 2.40 | 1.89 |
| | 1½ | 7.75 | 4.13 | 3.95 | 3.11*** |
| | 2 | 4.40 | 3.65 | .60 | <1.00 |
| | 2½ | 2.70 | 3.18 | −1.10 | <1.00 |

* *t* values are based upon the error variance of the analysis of variance in Table 1.
** Significant at the .02 level of confidence.
*** Significant at the .01 level of confidence.

These results imply that the number of perceptual-motor adjustments increases as the amount of prerest practice increases. Since late in learning more adjustments are present, more can be lost during the rest interval. However, with large amounts of prerest practice, the set responses should develop more habit strength, and therefore be more resistant to forgetting. At first blush, our results seem contrary to the concept of habit strength. However, although late in learning the reduction constant of the forgetting curve is less than it is early in learning, the absolute amount of set lost late in learning is greater because there is more present that can be lost. In other words, although the percentage of set lost late in learning is less than that lost early in learning, the amount of adjustments lost is greater because there is more present that can be lost.

An alternative explanation would be that the acquisition of set involves not only the learning of new responses, but also the elimination of reactions, i.e., the dropping out of superfluous responses. Our results suggest that even if this latter hypothesis is correct, the number of correct responses acquired must be greater than the number of incorrect responses lost. If the reverse were true, less warming-up activity would have been necessary as the amount of prerest practice increased.

The importance of the set responses is also revealed in the groups that had no warming-up activity. Figure 1 indicates that retention for those groups did not vary with different amounts of prerest practice. To explain these results it is advantageous to assume that $I_R$ must have increased with learning. The dissipation of this greater inhibition offset the greater loss in set that occurred when rest was introduced late in learning. That more set reactions were lost late in learning is strongly suggested by the remainder of the curve depicted in Fig. 1.

Our finding that reminiscence following zero warm-up does not vary with increasing practice, is not incompatible with the results obtained by Wasserman (11). Using the alphabet printing task and two degrees of motivation, he found a decrease in reminiscence late in learning with low motivation but not with high motivation. It is possible that the motivation in the present experiment was much closer to Wasserman's high-motivation groups.

*Retention following massed and distributed warming-up activity.*—In Fig. 3 are shown the mean increases in score for the groups having either

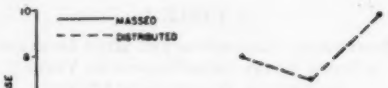


FIG. 3. The mean increase in score on the trial following various amounts of massed and distributed warming-up activity after 15 trials of prerest practice

massed or distributed warming-up activity. Results of the relevant groups of Part I are included in the curve for the massed warming-up activity. It is evident that the curve reaches its peak following 1 min. of massed warming-up activity, decreases beyond 1½ min. of warming-up activity, and then levels off following 2½ min. of warming-up activity. Apparently, warming-up activity in excess of a certain amount resulted in decreased performance. The fact that progressively increasing amounts of warming-up activity beyond a critical level does not result in progressively inferior retention scores needs explanation.

Apparently, after a certain amount of warming-up activity most of the perceptual-motor adjustments are reinstated. After this critical level has been reached, the effects of further warming-up activity are offset by the accumulation of $I_R$. Our results suggest that after 2½ min. of work on the intervening task, the ratio between the two variables ($I_R$ and reinstatement of set) becomes constant.

Figure 3 also suggests that after a certain amount of warming-up

activity, performance following rest decreases. (This trend is shown statistically in Table 2.) Apparently, immediately after the optimum point has been reached, the detrimental effects of $I_R$ are greater than the benefits resulting from reinstatement of set. Hence, before the curve levels off, the amount of $I_R$ accumulated and the amount of set reinstated is weighted in favor of $I_R$.

Reference to Fig. 3 indicates that the performance of the groups having distributed warming-up activity is superior to the performance of those groups having massed warming-up activity. An analysis of variance showed the $F$ ratio for distribution of practice on the warming-up activity to be significant at the .001 level (Table 3). These results are analyzed further in Table 4 which shows the significance of the difference between the mean increase in score following massed and distributed warming-up activity. It is evident that for 3, 4, and 5 min. of practice on the intervening task, distribution yielded a significantly greater postrest performance.

TABLE 3

ANALYSIS OF VARIANCE OF THE MEAN INCREASE IN SCORE ON THE TRIAL FOLLOWING VARIOUS AMOUNTS OF MASSED AND DISTRIBUTED WARMING-UP ACTIVITY AFTER 15 PREREST TRIALS

| Source of Variance | Sum of Squares | *df* | *F* |
|---|---|---|---|
| Distribution | 672.13 | 1 | 48.39* |
| Warming-up activity | 66.45 | 2 | 2.39 |
| Interaction | 33.22 | 2 | 1.20 |
| Error | 1583 | 114 | |
| Total | 2354.80 | 119 | |

* Significant at the .001 level of confidence.

TABLE 4

MEAN INCREASE IN SCORE AFTER REST FOLLOWING MASSED AND DISTRIBUTED WARMING-UP ACTIVITY AFTER 15 TRIALS OF PREREST PRACTICE

| Amount of Warming-up Activity (*Min.*) | Massed | | Spaced | | *t** |
|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | |
| 3 | 3.75 | 3.97 | 7.90 | 3.18 | 3.52*** |
| 4 | 3.10 | 4.29 | 6.95 | 2.21 | 3.26** |
| 5 | 3.75 | 3.28 | 9.95 | 4.38 | 5.26*** |

* *t* values are based upon the error variance of the analysis of variance in Table 3.
** Significant at the .01 level of confidence.
*** Significant at the .001 level of confidence.

It was indicated earlier that distributed practice on the warming-up activity should prevent the accumulation of $I_R$ for this set-maintaining activity. The reactive inhibition which is built up with each trial fades during the rest interval. Hence, performance on the main task cannot be depressed by $I_R$.

Figure 3 also reveals that increasing amounts of distributed practice on the intervening task tend to produce progressively greater increments in performance without any observable leveling off. Possibly all the perceptual-motor adjustments lost during rest had not been completely reinstated. Even if this explanation is correct, the marked rise in performance which is found with increasing amounts of distributed warming-up activity would not be expected.

However, another factor may have been operating. Ammons (1) suggests that the warming-up effect decreases as the number of rest intervals increases. In the groups having distributed warming-up activity the *Ss* had had practice in shifting from rest to work. This may explain the increase in retention which occurred with

increasing amount of distributed warming-up activity.

In general, it is evident that the results of this study have substantiated the major hypothesis and the related predictions, and thus have lent support to the theoretical basis of the experiment.

## Summary

Twenty-four groups of 20 *Ss* were given the alphabet printing task in order to test the following hypothesis: Warming-up activity aids retention as long as the benefits accruing from the reinstatements of set are greater than the detrimental effects of the accumulation of additional $I_R$.

A 10-min. rest pause was introduced after 5, 15, and 30 trials of prerest practice. In each of these conditions *Ss* were given 0, ½, 1, 1½, 2, and 2½ min. of warming-up activity, consisting of making tallies and crossbars. Six other groups having 15 trials of prerest practice received 3, 4, and 5 min. of either massed or distributed warming-up activity. The *Ss* in all the groups received 36 trials of 30 sec. each. The results and conclusions may be summarized as follows:

1. The *Ss* who received warming-up activity were significantly superior in performance on the trial following rest to those who had rest only, with no warming-up activity.

2. The amount of warming-up activity necessary to reinstate set was found to be an increasing function of the amount of prerest practice. This was interpreted as indicating that the number of perceptual-motor adjustments acquired during learning increased with increasing amounts of prerest practice.

3. With increasing amounts of massed warming-up activity, there was a tendency for retention to increase up to a point and then decline. This decrease indicates that the $I_R$ built up during the practice on the warming-up activity summated with the $I_R$ accumulated during the practice on the main learning task.

4. Increasing amounts of massed warming-up activity beyond a certain point did not result in a further decrease in performance. It was suggested that beyond a certain amount of warming-up activity, the ratio between the amount of set reinstated and the amount of $I_R$ accumulated becomes constant.

5. Because the accumulation of additional $I_R$ was prevented, distributed warming-up activity resulted in greater increments in performance than did massed warming-up activity.

6. Increasing amounts of distributed warming-up activity resulted in progressively greater increments in performance. The possibility was considered that the adoption of a particular set may have been facilitated by the constant shifting from working to resting with the distributed warming-up activity.

(Manuscript received November 29, 1951)

## References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
5. Irion, A. L. Retention as a function of amount of pre-recall warming-up. *Amer. Psychologist*, 1949, **4**, 219–220.

6. KIENTZLE, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.
7. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
8. RUCH, F. L., & WARREN, N. *Working with psychology*. Chicago: Scott Foresman, 1941.
9. SILVER, R. J. The effect of the amount and distribution of warming-up activity on retention in motor learning. Unpublished Ph.D. thesis, New York Univer., 1951.
10. SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1946.
11. WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potention in motor learning. *J. exp. Psychol.* 1951, **42**, 162–172.

# DECREMENTS AND RECOVERY FROM DECREMENTS IN A SIMPLE WORK TASK WITH VARIATION IN FORCE REQUIREMENTS AT DIFFERENT STAGES OF PRACTICE [1]

EDWARD A. BILODEAU

*Perceptual and Motor Skills Research Laboratory, Human Resources Research Center*

Almost all of the recent research in motor behavior has dealt with motor tasks in which *S* gradually acquires considerable proficiency. In the theoretical analysis of the data two different types of intervening variables have been postulated: one for the learning factor (${}_sH_R$) and one for the work factor ($I_R$). Typically, $I_R$ and ${}_sH_R$ vary together and are studied in controlled situations where the effects of one can usually be isolated from those of the other. It is equally desirable, however, to study one of these factors in situations where the other remains fairly constant and equal between conditions throughout practice.

It is also true that in the typical motor performance experiment the *n*th response is often unlike the first. This appears to be a serious shortcoming whenever $I_R$ is related to duration of the practice session or number of previous responses.

A device has been constructed for this study which presumably allows for equal ${}_sH_R$'s between conditions and fairly constant amounts of ${}_sH_R$ throughout practice. This has been accomplished by demanding a relatively simple response which can be assumed to be fairly well established in the response hierarchy. Further, decrements in performance have been induced through manipulations of the amount of force required to accomplish a simple motion rather than the much more customary procedure of varying the duration of the practice period (1).

Two different forces were required to turn a manual crank. One group practiced with one of these forces while a second group practiced with a second force. The third and fourth groups used the same forces as the controls but after each minute of practice the force required for handle-turning was changed. The two experimental groups differed from each other in that they never practiced with the same force during the same minute. In other words, they were out of phase. It was thus hoped that by systematic variation in the force required to move a resistant object the effects of force and locus of change in force could be isolated.

## METHOD

*Subjects.*—A total of 160 basic airmen trainees at Lackland Air Force Base was divided equally, but unsystematically, into four groups. No *S* had received prior practice with the apparatus.

*Apparatus.*—The basic apparatus consisted of a manual crank and a system allowing for controlled variation in the force necessary to turn the handle at any given rate of revolution.[2] The

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program and is described in (2). The opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force.

The author wishes to express his appreciation to 2nd Lt. Thomas G. Ferguson for constructing and calibrating the apparatus, and to Mr. Robert E. Morin, who supervised the statistical analysis of the data.

[2] A series of cranking experiments has been reported by Reed (4) in which such variables as radius of crank handle, braking force, orientation

TABLE 1

HORSEPOWER REQUIREMENTS FOR LOADS 0 AND 4 AT SEVERAL REPRESENTATIVE RATES OF ROTATION OF CRANKING DEVICE

| Load | Rate of Rotation (Rev./20 Sec.) | | | | |
|---|---|---|---|---|---|
| | 40 | 50 | 60 | 70 | 80 |
| 0 | .0095 | .0135 | .0180 | .0225 | .0280 |
| 4 | .0285 | .0410 | .0545 | .0690 | .0850 |

crank handle, with a turning diameter of 9 in., was mounted in the horizontal plane at the edge of a table. Upon the completion of each revolution of the handle, a counter recorded a score of one. The force required to turn the handle could be instantaneously varied by five steps, i.e., Loads 0, 1, 2, 3, and 4. The two extreme loads (0 and 4) were used in the present study. Horsepower (HP) values over a wide range of rate of revolution were calculated for Loads 0 and 4 by means of standard engineering calibration procedures. The HP requirements at representative rates of revolution (number of revolutions per 20 sec.) are presented in Table 1. This table makes evident the following properties of the crank system: (*a*) HP increases with increasing rate of rotation, (*b*) the difference in HP between the two loads increases with increasing rate of rotation, and (*c*) HP ratios for the two loads remain constant over all rates of rotation. For example, the ratio of Load 4 to Load 0 equals 3.06 regardless of the rate of rotation. It is also true that the ratio of the force for Load 4 to that for Load 0 equals 3.06 for any given rate of rotation.

*Procedure.*—The *S*s of all four groups, standing before the handle and using a standard grip, rotated the handle as fast as possible for an initial continuous practice period of 5 min. (initial practice). After a rest period of 40 sec., they practiced for an additional continuous practice period of 3 min. (final practice). Group 0 practiced with the minimum required force (Load 0); Group 4 practiced with the maximum force available (Load 4). The third and fourth groups, (Groups 0–4 and 4–0), alternated between these two loads after each minute of initial practice, Group 0–4 beginning with Load 0, and Group 4–0 beginning with Load 4. All four groups practiced *throughout* final practice under conditions identical with those obtaining during the last minute of initial practice, that is, Groups 0 and 0–4 practiced with Load 0, and Groups 4 and 4–0 practiced with Load 4.

No *S* was told about the durations of the practice periods nor instructed about the crank loads. The number of crank revolutions per 20-sec. periods was recorded.

## RESULTS

*Analysis of scores in terms of rate of rotation.* The mean number of revolutions per 20 sec. is plotted against successive 20-sec. periods for both initial and final practice in Fig. 1. The upper curve is for Load 0, the bottom curve is for Load 4. Both curves appear to have the same origin, affording some basis for the contention that initially there were equal habit strengths for both groups. With practice, however, there are marked decrements for both, the performance with the heavier load falling off more rapidly for the first minute. After the first minute there appear constant differences in output. The indication is that each is approaching a different minimum.

During the first minute, Groups 0 and 0–4 have worked with Load 0, Groups 4 and 4–0 have worked with Load 4. The mean difference between loads, evaluated by means of the within-groups error estimate, is significant beyond the 1% level ($t = 3.749$, with 156 *df*). When Group 4–0 is instantaneously shifted from Load 4 to Load 0 after 1 min. of practice there is a marked increment in scores but not quite to the level of the Load 0 control (difference significant at the 6% level with $t = 1.927$ and 156 *df*). When Group 0–4 is shifted from Load 0 to Load 4 there is a pronounced decrement in output to a level somewhat below the Load 4 control, but this new level in rate of responding does not differ significantly from the Load 4 control ($t = 1.648$, with 156 *df*). After each successive shift each group nearly

of the crank, height of the crank as well as handedness, direction of rotation, bodily position, and fatigue are discussed.

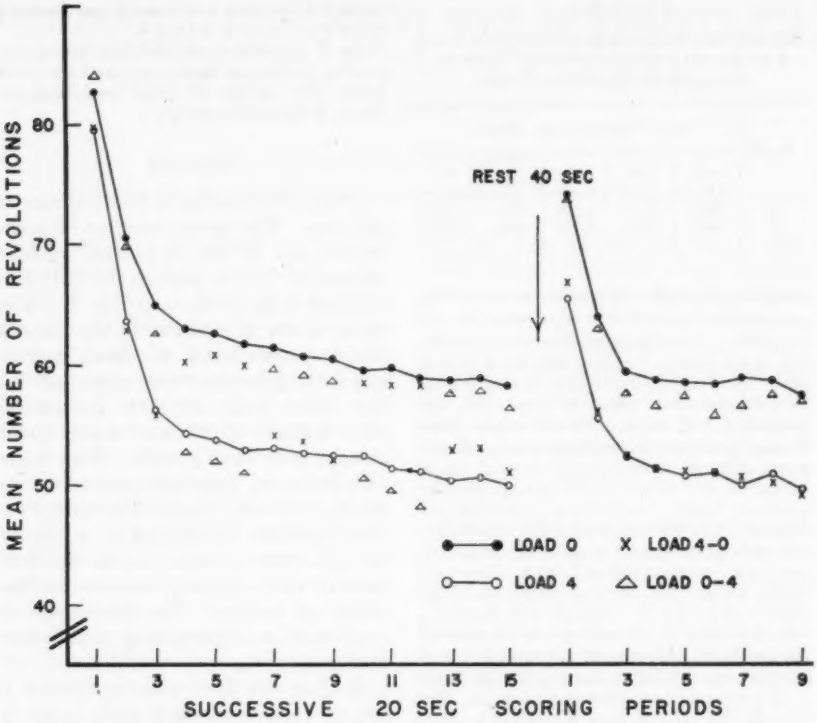


FIG. 1. Mean number of revolutions per 20 sec. is plotted against successive 20-sec. scoring periods

approximates its control. Further, the mean number of crank turns during the entire 5-min. period for Group 0–4 almost equals that for Group 4–0, and each is nearly equal to the average output of Groups 0 and 4. The mean totals for Groups 0, 4, 0–4, and 4–0 are 62.86, 54.88, 58.06, and 58.04, respectively.

After a rest period of 40 sec. each group shows considerable recovery, but none recovers to the initial level of responding. Numerically, each of the four groups recovered nearly identical amounts, the difference being far from significant by the $F$ test ($F = .38$, with 3 and 156 $df$).

The difference between Groups 0 and 4 after rest also remains fairly constant and is approximately equal to that of prerest practice. Group 4–0, practicing throughout final practice with Load 4, is remarkably similar to Group 4 in mean scores whereas Group 0–4, working with Load 0, is much like Group 0. Evidently, the alternation between forces during initial practice had little differential effect upon performance during final practice.

*Analysis of scores in terms of horsepower.*—When number of revolutions per 20-sec. periods is converted to an HP scale, output during both practice periods shows a direct relationship to load rather than the inverse relationship evident in terms of number of

revolutions. Further, a measure of HP recovery during rest was calculated by converting the last score of initial practice and the first score of final practice to HP and obtaining the difference. An average of .0075 and .0080 HP was recovered by Groups 0 and 0–4, while .0220 HP was recovered by both Groups 4 and 4–0. In other words, the two groups with the slower rate of rotation during the terminal period of initial practice recovered the most in terms of HP.[3]

*Analysis of decrement for Groups 0 and 4.*—An extensive analysis of the decrements for the individuals of the upper and lower quartiles has been published elsewhere (2). This analysis tended to show that those groups beginning a practice period at a level of responding greater than that for other groups exhibited greater decrements. The results were interpreted as indicating that the magnitude of response decrement at any time during continuous practice is some increasing function of the magnitude of the initial level of responding, $({}_{S}\bar{E}_{R})$.

## Discussion

After the first minute of practice the difference between the response rates of the two nonalternating control groups was nearly constant; with further practice they showed parallel decrement; with rest both groups spontaneously recovered equal amounts. The two alternating groups showed a marked tendency to approximate the rate of their respective controls whenever the same load was employed. This was true with instantaneous shifts in force treatment before rest and true also after rest when the groups might have been expected to differ because of the prior alternation treatment.

Thus, the data of both initial and final practice have indicated that rate of responding is a function of the prevailing load and of the duration of previous practice. The load obtaining during prior practice does not appear to be particularly effective in producing differential response rates in subsequent practice. This suggests that the output of the responding member represents an adjustment to the duration of previous practice and to the load prevailing in such a manner that the former load is almost irrelevant in determining present rate.

Each of the above findings is consistent with the hypothesis that the *S*s of each group are actually about equally fatigued. Such a conclusion is surprising to those who, perhaps, intuit that work against a heavy load is more fatiguing than work against a light load. This interpretation fails to recognize that *S* is free to choose his own pace. If it is granted that the present groups were about equally fatigued in spite of working against different loads and different time distributions of these loads, then it must be granted that *S*s choose their own pace according to the load condition. The hypothesis here is that there is a similar state of tolerance for fatigue from group to group. Given a single state of tolerance for fatigue, differential rates are established as a function of duration of previous practice and present load. Differences in rate attributable to load are such that load is not relevant for predicting subsequent performance level. In order to maintain similar levels of fatigue from group to group there must be a difference in rate of responding between groups working against different loads. In simplest terms, it is as if any *S* had only so much to give and that

[3] A second experiment now in preparation will deal with spontaneous recovery in greater detail.

rate (pacing) represents an adjustment to this state of affairs.

It is interesting to contrast the foregoing interpretation of the results with the theoretical assumption underlying the choice of a cranking task in the first place. The task was originally designed so as to provide a situation eliciting equal ${}_{S}H_{R}$ from load to load in order that the learning factor would not be confounded with the effects of $I_R$. Thus, it is reasonable to venture that under the conditions of the present study, the habit, drive, and decrement terms of the equation for determining momentary effective behavior potential are equal from load to load.

It will be recalled that HP for the heavy load group was greater than for the light load group. This is, of course, a statement which is relevant to work which is done against the external environment. In terms of energy expenditure, however, conversion of revolutions to HP neglects the weight of *S*'s arm. If the weight of the faster moving arm of the light load groups were taken into account, it seems reasonable to suppose that the HP expended by this group would be similar to that expended by the heavy load group.

The interpretation of the present data has in effect suggested that the phenomena of massing-spacing of practice in motor learning studies be reexamined in regards to possible pacing effects. It could be, for example, that at least some of the difference in rate of responding results not so much from differential fatigue as from differential pacing appropriate to the distribution stimuli.

## Summary and Conclusions

The *S*s ($N = 160$) practiced rotating a crank handle as fast as possible for 5 min. with either a light or heavy loading of the crank. One group practiced with the light load, a second practiced with the heavy, and the third and fourth groups alternated between loads after each minute of practice but were out of phase. After a brief rest all groups practiced for 3 min. with the loading used just before the rest.

During the first minute of practice, rate of responding dropped more rapidly with the heavy load than with the light load. Thereafter, differences between the loads remained fairly constant. After a brief rest both load groups recovered equally in terms of rate, but the heavy load group recovered more in terms of HP.

When the two groups which had practiced under one load condition were instantaneously shifted to the other, rate output was approximately equal to that of the nonshifting groups. Furthermore, after rest both alternating groups spontaneously recovered to the level of their nonalternating controls as if unaffected by the alternation treatment of initial practice.

(Manuscript received December 14, 1951)

## References

1. Bilodeau, E. A. Performance in a simple motor task before and after a single rest period. Unpublished Ph.D. thesis, State Univer. of Iowa, 1950.
2. Bilodeau, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. USAF Air Training Command Human Resources Research Center *Research Bulletin* 51–34, December 1951.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Reed, J. D. Factors influencing rotary performance. *J. Psychol.*, 1949, **28**, 65–92.

# LEVEL OF REPETITION IN THE "SPREAD OF EFFECT"[1]

FRED D. SHEFFIELD AND WILLIAM O. JENKINS
*Yale University* *Harvard University*

In an earlier study by the writers (3) it was shown that guessing habits are a major factor in accounting for the "spread of effect." The original study has been followed by a number of investigations (1, 2, 3, 4, 7, 8, 9), most of which have provided ample confirmation of the guessing-sequence interpretation. It appears that Thorndike's "spread-of-effect" phenomenon, which he considered a proof of the automatic action of reward, is an artifact due chiefly to *S*'s tendency to guess numbers systematically rather than randomly. The operation of this artifact is very clearly seen in a study (7) that demonstrated a spread gradient under circumstances in which numbers were guessed without either reward or a learning task. In this study, however, while a clear-cut gradient was found, the over-all level of repetition was considerably lower than the levels typically obtained in Thorndike's investigations involving a comparable number of choices (10).

The gradients obtained without reward or learning (7) are compared in Table 1 with the average of three appropriately selected experiments by Thorndike (10, Exp. F, H, and I).

Thus, in the position immediately after reward, Thorndike obtained around 20% repetition when using numbers from 1 to 10. But in the experiment without reward or learning, a value of only 13.3% was found in this position. The purpose of the present experiments was to investigate some of the factors that might account for this discrepancy. A more general purpose was to provide information on the relative importance of the various factors that influence the level of repetition in the "spread-of-effect" phenomenon.

A factor thought to be of importance in accounting for the difference in levels of repetition in the Thorndike situation and in that without learning was the presumed distracting effects of the learning task. In the study of spread without reward or learning (7), the *S*s thought they were participating in an ESP experiment and that the task was to reproduce, through "clairvoyance," designated portions of a printed table of random numbers. Specific instruction as to the nature of randomness was given. There was no learning task and *S* could concentrate on guessing numbers at random. The *S* wrote his numbers on a prepared form and could see his prior guesses. In the typical Thorndike experiment, on the other hand, *S*'s task is to learn word-number pairs—a situation that could distract him from carrying out instructions to guess numbers "at random." Thorndike's *S*'s usually guessed aloud and could not keep track of prior guesses visually.

TABLE 1

COMPARISON OF EXPERIMENTS

| Experiments | Position in after Gradient | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Without reward or learning (7) | 13.3 | 11.4 | 11.2 | 10.6 | 10.2 |
| Thorndike (10) | 20.4 | 19.3 | 14.5 | 17.3 | 15.5 |

[1] This research was facilitated by the Laboratory of Social Relations, Harvard University, and Yale College. The writers are grateful to Dr. F. Mosteller for valuable comments on the manuscript and to Mr. James Stephens who ran the individual *S*s as a Bursary assistant financed by Yale College.

Thus, if *S* has systematic tendencies in number guessing, these might be maximized in the Thorndike situation with attention directed to a learning task, and minimized with attention directed solely to reproducing a table of random numbers. This hypothesis had an a priori reasonableness, and previous findings (3) had indicated that the level of repetition was higher under conditions of distraction. To test this hypothesis, guessing sequences were obtained under two conditions: (*a*) no distraction, and (*b*) with the distraction of an irrelevant learning task going on at the same time numbers were guessed. An attempt was made to duplicate the distracting aspect of Thorndike's task under conditions which precluded the action of reward as a source of "spread of effect," but allowed the operation of the guessing-sequence factor. In addition to this major comparison, evidence was collected on the effect of other factors which could raise the level of repetition in the Thorndike situation.

## Method

Two independent sets of experiments were carried out, one using an individual testing procedure at Yale, the other employing group testing at Harvard. The *S*s were undergraduate students. The basic design consisted of comparing the performance under experimental conditions, in which *S*s guessed numbers while simultaneously learning verbal material, with the performance under control conditions, where numbers were guessed without the distraction of any learning task.

The distracting task under the experimental conditions was paired-associates learning of three-letter meaningful words. These were randomly selected and paired from an available exhaustive list (6). In all experiments, *S*s were instructed to guess a number from 1 to 10 at random as soon as the stimulus term of a pair was exposed. In some experiments the anticipation procedure was used; in others, tests with only stimulus terms followed presentation of the pairs. In all cases, both in acquisition and testing, *S* was required to guess his number immediately after the presentation of each stimulus term. With the individual procedure *S* called out his number and *E* recorded it; with the group procedure *S* wrote his number on an answer sheet and immediately covered it with his hand. With one exception to be noted later, the exposure rate used throughout all experiments was 4 sec.

Under the experimental conditions, *S*s were told that the purpose was to test the effect of distraction on learning. The requirement of guessing a number was explained to *S* as an attempt to provide a distraction which would interfere with the learning process. Under the control conditions, *S* was asked to guess numbers from 1 to 10 in response to a 4-sec. signal, with no rationale offered other than an appeal to his faith in the scientific importance of the study.

With the individual testing procedure, a modification of the Hull exposure apparatus was used. In the main experiment ten paired associates were employed, and these were varied in six random orders of presentation, using the standard anticipation procedure. The roll of six random changes was run through twice, yielding a total of 120 number guesses per *S*. The same *S*s ($N = 22$) served under both the experimental and control conditions. Experimental conditions were always run first. The decision not to randomize the order of the experimental and control conditions was based partly on the assumption that a control was provided by the previous study referrred to (7). A further important consideration was the possibility that running *S*s first under the control conditions might lower their motivation when they subsequently learned under "distraction." Two auxilliary groups ($N = 15$ and $N = 9$) were run with nonanticipation techniques to be described later.

In the group-testing procedure, the experimental condition involved 60 paired associates printed on flash cards. Both members of a given pair were exhibited simultaneously in the training trial. The nonanticipation method of testing was used; the test trials consisted of presenting only the stimulus terms in a different random order from that employed in the training trial. Thirty-two *S*s were run under these conditions; 52 additional *S*s were run under the control conditions of number guessing only. Under both experimental and control conditions, 120 number guesses per *S* were obtained.

An additional 11 *S*s were run with the group procedure using the same order of stimulus-term presentation on the test as had been used on the training trial. This group provided evidence on the effect of the fixed-list factor, present in all of Thorndike's studies and a possible source of greater repetition through the action of word-number associations.

The scoring procedure in all experiments was

similar to that described previously (7). The *S*'s total series of guessed numbers were divided into sequential subsets (to be described later in connection with each experiment). Successive subsets were aligned in such a way that *S*'s guessed numbers in the same ordinal position within subsets could be compared. Each repetition of the same guess in the same ordinal position in two successive subsets were scored as a *repeat*, and the presence or absence of other repeats in the five following positions was determined. This scoring method parallels that of Thorndike (10); the only important difference is that "rewarded" responses in Thorndike's procedure are replaced by chance repetitions in the present experiments.

In applying significance tests the percentage repetition score of each *S* (at a given point on the gradient) was treated as a single observation. Thus *N* in each comparison is the number of *S*s involved rather than the total number of repetitions. Except where the contrary is noted, the 1% level of significance was accepted.

## Results of Main Experiments

*Individual testing results.*—In the main experiment with individual testing, each of the 12 runs through the list of ten pairs was considered a trial, (i.e., a "subset"), and each such trial was compared with the succeeding one to obtain the "after" gradient. The data obtained are shown in Table 2.

It can be seen that the expectation of a higher level of repetition with a distracting learning task is not borne out. The difference is for the most part in the opposite direction, although not significantly so at any point on the gradient. The *p* value for the largest difference in Table 2 (Position 3) was .21. However, both gradients are significantly higher in level than the one obtained in the previous study (7). Thus, whereas the previous result obtained a level of only about 13% repetition in the first position, the present finding gives levels of about 17% and 19%, respectively, for the experimental and control conditions.

TABLE 2

Gradient of Repetition Following Chance Repeats: Individual Testing with the Anticipation Procedure and Each *S* Serving as His Own Control (22 *S*s)

| Condition | Position after Chance Repeat | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Mean per cent repetition with learning task | 17.2 | 14.7 | 11.7 | 11.2 | 14.7 |
| Mean per cent repetition without task | 19.2 | 16.9 | 15.6 | 11.7 | 12.7 |

TABLE 3

Gradient of Repetition Following Chance Repeats: Group Testing with the Nonanticipation Procedure

| Condition | Position after Chance Repeat | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Mean per cent repetition with learning task (32 *S*s) | 16.9 | 13.7 | 13.1 | 13.1 | 10.0 |
| Mean per cent repetition without task (52 *S*s) | 15.4 | 14.4 | 10.9 | 11.9 | 7.5 |

It would appear that omitting instructions about the nature of random numbers and eliminating the chance to keep track of guesses visually, raised the level of repetition to values closer to Thorndike's, but that the addition of a distracting task had no further effect.

*Group-testing results.*—In the main group experiments *S*s wrote their numbers on forms having four columns of 30 spaces each. Gradients were obtained by comparing each column with the succeeding one, as previously described. The results are shown in Table 3.

As in the individual test results, the initial hypothesis is not borne out to a statistically significant degree. What differences were obtained are in the direction expected from the task-dis-

traction hypothesis, but the differences are not significant either over-all or at individual points. The smallest $p$ value for any of the differences between experimental and control conditions shown in Table 3 is .20. As in the individual testing results, however, the levels of repetition are appreciably higher in positions immediately after chance repeats than in the earlier group study (7).

## Results of Amplifying Experiments

*Anticipation versus nonanticipation in individual testing.*—From the guessing-sequence hypothesis it is expected that if the task is too distracting, guessing sequences will disappear. Thus the hypothesis depends on each guess being a stimulus which influences the next guess as a response. If a very distracting task intervenes between each guess, it might render the preceding guess an inconsequential aspect of the stimulus pattern at the time the succeeding guess is made.

It was thought that the anticipation method used in individual tests might involve a greater degree of distraction than Thorndike's procedure. Thus it might be expected that a nonanticipation procedure would yield levels of repetition more closely approaching Thorndike's. To test this hypothesis, 15 *S*s were run with a nonanticipation individual procedure comparable to that used in group testing.

The mean per cent repetition from positions one through five was: 15.8, 18.0, 12.2, 12.7, and 9.3. By comparing these figures with the top line of Table 2 it can be seen that the hypothesis is not supported; the nonanticipation method did not generally increase the level of repetition over that obtained with the anticipation procedure.

*Acquisition versus recall in individual testing.*—With the anticipation procedure, *S*'s task is a mixture of trying to learn each pair that has not yet been mastered and to recall the correct response term before it is presented. In the Thorndike experiment, however, *S* does not have an opportunity to rehearse each pair—at least in the early part of learning where *S* has to find out what the pairs are by guessing. For the most part, as each stimulus is presented, *S* is attempting to recall any response which may have been called "right" on a previous trial or is attempting to recall what responses have been called "wrong" on previous trials. Thus it can be argued that the Thorndike situation is one in which the chief distraction is an attempt to recall rather than to rehearse.

In an attempt to separate the factors of rehearsal and recall that distractors which contribute to spread, a modified nonanticipation procedure was used with nine *S*s. The same type of list (i.e., ten pairs and six random changes) was used that had been used in the main individual experiment. However, instead of attempted anticipation when each stimulus term was shown, *S* alternately concentrated on memorizing the pairs (both terms shown simultaneously) without overt anticipation on odd-numbered trials and attempting recall—to stimulus terms only—on even-numbered trials. As in the main experiment the list was presented 12 times, giving six training trials with pairs presented alternated with six tests in which only the stimuli were presented.

Each *S* learned two different lists under the above conditions. The two lists were learned under different presentation rates, one with a 4-sec. cycle and the other with a 2-sec. cycle. Analysis consisted of comparing successive training trials and, separately, successive tests, with the same pro-

TABLE 4

"SPREAD" AROUND CHANCE REPEATS WHILE MEMORIZING (TRIALS) AND ATTEMPTING RECALL (TESTS) OF AN IRRELEVANT LEARNING TASK WITH EACH *S* SERVING AS HIS OWN CONTROL (9 *Ss*)

| Measure | Position after Chance Repeat | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Trials | 20.1 | 14.9 | 12.0 | 10.4 | 13.3 |
| Tests | 30.6 | 21.9 | 18.9 | 20.0 | 17.6 |

cedure as used in the main individual experiment. The results are summarized in Table 4. No consistent effect of the presentation rate was found, so the results are broken down only by "trials" versus "tests," an analysis which presumably pits the distracting effects of rehearsal against those of attempted recall, respectively.

It should be noted that the results are fairly unstable with only nine *Ss* even though each *S* learned two lists. However, while the "trial" results are roughly comparable with the findings in the main individual experiment, the "test" results show a significantly higher level of repetition at all points on the gradients. The results indicate that guessing sequences are more likely to operate when *S*'s symbolic activity is occupied in attempting verbal recall of the correct response than when both members of a pair are shown and *S* is silently attempting to memorize these.

*Use of a fixed list in group testing.*—A feature of Thorndike's experiments (10) which could increase the over-all level of the after gradient was his use of fixed orders of stimuli in successive presentations of lists. In all the present experiments described thus far *mixed* lists were used. That is, the order of the stimuli was varied on successive acquisitions and test trials in individual testing and was uncorrelated in training and test in the group testing. This procedure precludes the possibility that initial word-number associations might raise the level of repetition in the after-gradient. Thus, for example, if *S* has an initial tendency always to guess "two" when the stimulus is "leg" he will tend to repeat himself in that same position if the list is fixed. To the extent that such initial associations operate, the entire level of repetition of "wrongs" is raised. Moreover, if an association is formed during the experiment between stimulus terms and the wrong response number guessed, any tendency for the association to persist will show up as an increased repetition in the same position of a fixed list.

A check on the effect of fixed order was provided by an auxiliary group experiment ($N = 11$) in which a fixed order of stimuli was used in both training and testing. Otherwise the procedure was exactly the same as the main group experiment (with learning task), and the same list of pairs was used.

The mean per cent repetition from Position 1 through 5 was: 20.7, 18.8, 14.3, 13.2, and 7.4. By comparison of these results with the results for the task groups in Table 3, it is apparent that the level of repetition is higher with the fixed than with the mixed list. The differences at the first and second points are significant at the 2% level of confidence. The findings may be interpreted as indicating that an important factor explaining Thorndike's level of repetition in his typical-spread studies was his consistent use of fixed-stimulus orders.

## DISCUSSION

The main import of the present results is that guessing sequences seem sufficient to account for Thorndike's spread phenomenon. In the earlier

study (7), the nature of random numbers had been carefully explained and *Ss* had an opportunity to keep track of their sequences visually. Even under these conditions a spread gradient was found, but with a low level of repetition. In the present studies the most comparable procedure *without* either instructions on randomness or opportunity to keep track of sequences is the control condition in the main group experiment. As can be seen in Table 3 the level of repetition was raised by eliminating these factors. Also, when Thorndike's fixed-list factor was included, the level of repetition was further raised. Moreover, when the task was made—as in Thorndike's situation—primarily a matter of recall rather than acquisition, an appreciable increase in repetion was obtained (Table 4). It would appear, therefore, that the more closely the factors operating in the Thorndike situation are approximated, the more nearly Thorndike's levels of repetition are achieved. In none of the present studies did any presumed "spread of effect" of reward have an opportunity to generate the spread gradient; in all cases the analysis was based on "spread of effect" following chance repeats. A result very consistent with the present findings has recently been reported by Fagan and North (1) who ran *Ss* individually, apparently did not train their *Ss* in the nature of randomness, and prevented their *Ss* from keeping track of their number sequences visually. As in the present studies, their gradients show a higher level of repetition than in the original study (7) of "spread of effect" without reward or learning.

It should be noted that in all the present studies, and in previous comparable ones (1, 7), the analysis procedure involves 100% repetition in the "zero" position. In Thorndike's studies the percentage repetition at the zero ("rewarded") position was less than this figure. Thus, the present studies, in attempting to reproduce Thorndike's gradients, might yield an even higher level of repetition than in Thorndike's studies because of the greater frequency of repetition at the zero point. However, in the experiment that most closely reproduced Thorndike's causal factors (see Table 4), the levels are as high as those reported by Muenzinger and Dove (5) who experimentally induced 100% repetition at rewarded points by teaching certain word-number associations in advance.

A final word should be said about the relation between the present results and the "probability bias" hypothesis offered by Smith (8). Subjects have a variety of habits in guessing numbers, some idiosyncratic and some fairly general for the population (2, 3, 8). Smith's "probability bias" factor—the tendency for *Ss* to avoid repeating the same guess two or more times in a row—is a very common tendency noted earlier (3). In guessing numbers from 1 to 10, however, it would account only for a gradient that starts with 11% repeats in Position 1 after and falls off to 10% (chance) in Position 5 after. But even this factor is subsumed under the hypothesis of habitual guessing sequences as the source of spread. As Smith notes in his article, a more general statement of "probability bias" would include any tendency on the part of *S* to fail to guess at random. Presumably any failure to guess at random is caused by nonrandom *habits* of guessing. The factor stressed by Smith is operative with almost every *S*, but is only a small contributor to the gradient or to the over-all level when ten choices are used. The present results attempt to get at the more important factors by duplicating everything in Thorndike's procedure except the automatic strengthening effect of a reward on adjacent connections. The results give further support to the interpretation that Thorndike's phenomenon is an artifact of guessing habits.

It should be noted that the distraction factor investigated in the present studies and previously shown (3) to appreciably raise the level of the spread gradient could, under certain circumstances, *lower* the gradient by interfering with guessing sequences. According to the guessing-sequence interpretation, the preceding guess functions as a stimulus in influencing the following guess, depending on *S*'s individual number-guessing habits. A distraction will *facilitate* sheer habit if the distraction interferes with *S*'s carrying out instructions to reproduce random sequences rather than such habitual sequences as counting up or down. A highly effective distraction, however, would also interfere with the tendency of a preceding guess, as a stimulus, to influence a succeeding guess as a response. Thus,

under certain conditions of distraction, all guessing habits might be eliminated as determiners of S's sequences. Apparently, the distractions in the main experiments were somewhere between these two extremes. In comparing the distractions of rehearsal and recall as tasks, however, it would appear that the recall task is a more optimal condition for the operation of guessing habits.

The fundamental question asked by the present experiments is, of course, whether the guessing-sequence interpretation (3) of Thorndike's "spread" effect can account entirely for the phenomenon. It had been shown (7) that pure guessing of numbers produced a Thorndike gradient without reward or learning, but at a much lower over-all level than in the typical Thorndike experiment. The present studies can be regarded as an attempt to approach Thorndike's conditions while at the same time keeping the number-guessing task and the rote-learning task completely divorced from each other. The experiments were successful in raising the level of repetition to the values reported by Thorndike, and to even higher levels, when his conditions were reinterpreted from the standpoint of the guessing-sequence hypothesis. The chief conclusion, therefore, is that Thorndike's explanation of his results called attention to the wrong causal factors present in his experimental situation.

## Summary

A number of experiments were performed to study factors that influence the level of repetition in the "spread-of-effect" phenomenon. The experiments tested implications of the guessing-sequence hypothesis. Some groups guessed numbers while concentrating on an irrelevant learning task; other groups merely guessed numbers without any distracting task.

The level of repetition was raised by omitting instructions on randomness and opportunity to keep track of sequences. Moreover, Thorndike's fixed order of presentation of stimulus terms raised the level still further. Finally, when the task emphasized recall rather than acquisition of the learning material, an additional increase in repetition was found that considerably exceeded the levels obtained by Thorndike.

The conclusion is drawn that the more closely Thorndike's situation is approximated—as reinterpreted from the standpoint of the guessing-sequence hypothesis—the higher the level of repetition even when reward cannot generate the gradients. It appears that guessing sequences are sufficient to account for Thorndike's spread phenomenon.

(Manuscript received December 14, 1951)

## References

1. Fagan, C. A., & North, A. J. A verification of the guessing-sequence hypothesis about spread of effect. *J. exp. Psychol.*, 1951, **41**, 349–351.
2. Jenkins, W. O., & Cunningham, L. M. The guessing-sequence hypothesis, the "spread of effect," and number-guessing habits. *J. exp. Psychol.*, 1941, **39**, 158–168.
3. Jenkins, W. O., & Sheffield, F. D. Rehearsal and guessing habits as sources of the "spread of effect." *J. exp. Psychol.*, 1946, **36**, 316–330.
4. Marx, M. H., & Bunch, M. E. New gradients of error reinforcement in multiple-choice human learning. *J. exp. Psychol.*, 1951, **41**, 93–104.
5. Muenzinger, K. F., & Dove, C. C. Serial learning: I. Gradients of uniformity and variability produced by success and failure of single responses. *J. gen. Psychol.*, 1937, **16**, 403–413.
6. Sheffield, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished Ph.D. thesis, Yale Univer., 1946.
7. Sheffield, F. D. "Spread of effect" without reward or learning. *J. exp. Psychol.*, 1949, **39**, 575–579.
8. Smith, M. H., Jr. Spread of effect is the spurious result of non-random response tendencies. *J. exp. Psychol.*, 1949, **39**, 355–368.
9. Taylor, D. W. Spread of effect in one-trial learning in relation to the guessing-sequence hypothesis. *Amer. Psychologist*, 1947, **2**, 313 (Abstract).
10. Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contr. Educ.*, 1935, No. 580.

# MASSING AND SPACING PHENOMENA AS FUNCTIONS OF PROLONGED AND EXTENDED PRACTICE [1]

EDWARD A. BILODEAU

*Perceptual and Motor Skills Research Laboratory, Human Resources Research Center*

The effects of distribution of practice over periods of several days have hardly been investigated. The lack of data is particularly apparent within the area of motor behavior.

Since the concepts of work and learning have been described as essential for the explanation of motor performance (**4, 5**), it has been desirable to study the role of each in situations where the effect of one can be isolated from that of the other or where the effect of one is negligible. The decremental effects of practice have often been held to a minimum by extreme spacing of practice, while the effects of learning can be held to a minimum by devising a situation where the appropriate responses are readily available as *S* begins practice.

In the present investigation work has been varied by the spacing of practice and learning has been minimized by demanding a relatively simple though effortful response. Further, the use of a simple response has given greater assurance that terminal responses are replicates of responses initiated early in practice. This desideratum is not always obtained in studies where learning is assumed to have taken place.

The purpose of this study was to relate (*a*) level of responding or rate of response early and late within a practice session, and (*b*) amounts of spontaneous recovery for relatively short and long rest periods, to the number of previous days of practice and to the manner of distribution of prior practice under conditions where the response is relatively simple but effortful.

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force.

## METHOD

*Subjects.*—The *S*s of this experiment were 52 airmen at Lackland Air Force Base who had just completed basic training and were awaiting assignment. None of these *S*s had received prior practice with the test.

*Apparatus.*—The apparatus was the manual crank described in previous publications (**1, 2**). In brief this apparatus consists of a crank handle (with a turning diameter of 9 in.) mounted in the horizontal plane at the edge of a table top and a tachometer generator for exerting a variable braking force upon the crank shaft. The properties of the internal system are such that the force required to turn the handle at any given rate of rotation is known. Although five different force variations were available, only one was used for the present investigation. The one employed has been previously described as Load 4 (**3**). With this load .0410, .0545, .0690, and .0850 horsepower, for example, are required at 50, 60, 70, or 80 revolutions per 20 sec., respectively. Any of these rates of revolution appears to be quite strenuous and leads to reports of great muscle fatigue. The number of revolutions per 20-sec. periods was recorded.

*Procedure.*—The *S*s were divided equally and at random into two groups and instructed how to stand before the handle, to use the standard grip, and to rotate the handle as fast as possible throughout the practice session.

For both groups, practice consisted of 8 min. of cranking on each of ten consecutive weekdays. Group 8, the massed-practice group, practiced without rest during the entire 8 min. each day. Group 4–4–4, the spaced-practice group, practiced continuously for 4 min., rested for 4 min., and then finished practice with a second continuous 4-min. practice period.

The *S*s of the two groups were run alternately, that is, an *S* of the massed group always followed

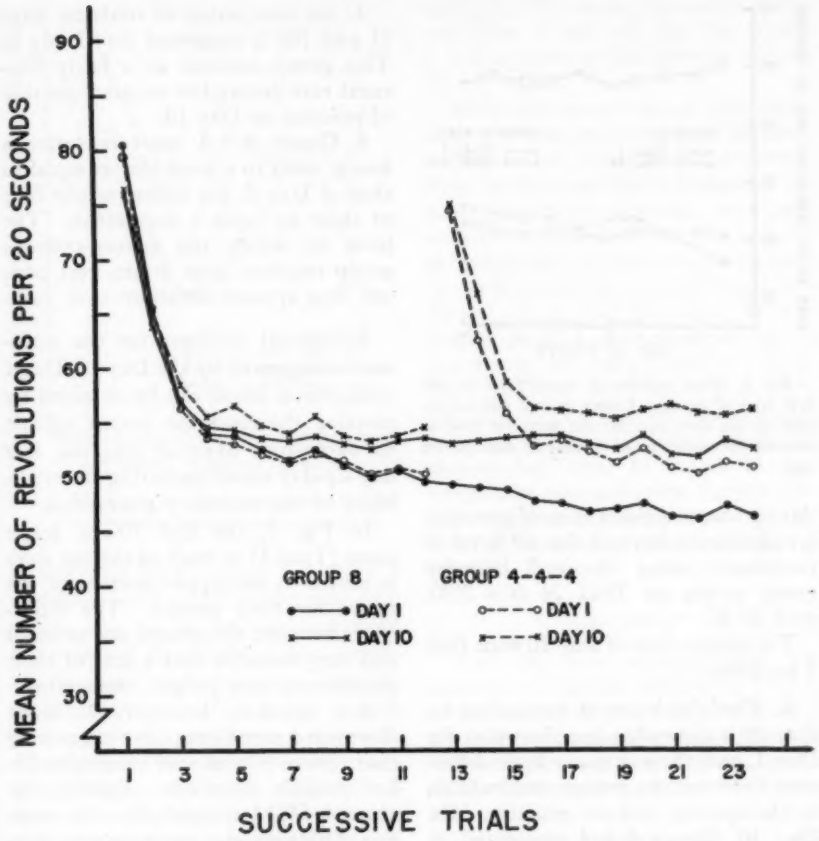


FIG. 1. Mean number of revolutions per successive 20-sec. trials on Days 1 and 10 with massed-practice (Group 8) and spaced-practice (Group 4-4-4)

an *S* of the spaced group. Each *S* was run at approximately the same hour each day. Before each session, all *S*s were reminded to crank as fast as possible between the "go" and "stop" signals. This reminder was not repeated for the spaced-practice group, however, during the interpolated rest period. Two *S*s from Group 8 were lost because of scheduling difficulties.

## RESULTS

The trial means for Day 1 and Day 10 for each group are presented in Fig. 1.[2] The Day 1 means for the first 12 trials for both groups are remarkably similar, 4 min. of continuous practice being the experimental treatment ($t = .473$, with 48 *df* when computed with the scores of Trial 12). After Trial 12 and a rest of 4 min., Group 4-4-4 spontaneously recovered about 24 rev./20 sec., or 93.67% of the initial rate of responding. Group 8, continuing practice without rest, continues to decline throughout the remaining 12 trials. On Day 1 the spaced group remains above Group 8

[2] Detailed tables of data from this study are available in (2).

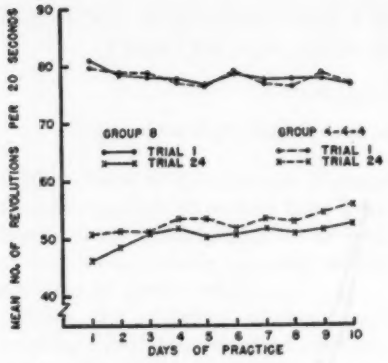


FIG. 2. Mean number of revolutions for the first trial of practice (upper pair of curves) on each of ten days and for the terminal trial of practice (bottom pair of curves) on each of ten days

during the terminal 4 min. of practice, a $t$ significant beyond the .01 level of confidence being obtained between group scores on Trial 24 ($t = 2.90$, with 48 $df$).

The comparison of Day 10 with Day 1 suggests:

1. The initial rate of responding on Day 10 is somewhat less than that for Day 1, and there appears little difference between the groups attributable to the spacing of prior practice. On Day 10, Group 4–4–4 responded at 96.03% of its Day 1 level and Group 8 responded at 94.98% of its Day 1 level.

2. After 4 min. of practice on Day 10 both groups were cranking faster than for the comparable point on Day 1, and there appears to be little difference between the groups. When the groups were compared by means of $t$ on the Trial 12 scores of Day 10, no significant difference was obtained ($t = .375$, $df = 48$). Pooling the data of the two groups, however, resulted in a $t$ for Day 1 versus Day 10 which was significant far beyond the .01 level ($t = 4.08$, $df = 48$).

3. An interaction of trials by days (1 and 10) is suggested for Group 8. This group cranked at a fairly constant rate during the terminal portion of practice on Day 10.

4. Group 4–4–4 recovers (after a 4-min. rest) to a level almost equal to that of Day 1, but subsequently fails to show as rapid a decrement. The level to which the spaced-practice group recovers after 4-min. rest periods thus appears invariant over days.

Additional evidence for the statements suggested by the Day 10–Day 1 comparison above can be obtained by plotting the condition means against 10 successive days of practice, any day-by-day trend increasing the credibility of the summary statements.

In Fig. 2, the first 20-sec. mean score (Trial 1) on each of the ten days is shown in the upper portion of the figure for both groups. The differences between the groups are so small and unsystematic that a test of their significance was judged unnecessary. There appears, however, a slight downward trend over days suggesting that recovery level over successive 24-hr. periods decreases slightly (by about 4–5%) but regularly. An analysis of between-day variance over day-by-subject variance for the pooled distributions yielded an $F$ significant beyond the .01 level of confidence, confirming the downward trend ($F = 4.57$, 9 and 441 $df$).

The Trial 24 means have been plotted in the lower portion of Fig. 2 for each condition. The upward trend of the data for both groups suggests that the terminal rate of responding is an increasing function of days of practice. This trend, too, appears to be real since the $F$ (between-day variance over day-by-subject variance) was significant beyond the .01 level for each group separately ($F = 9.26$ for Group

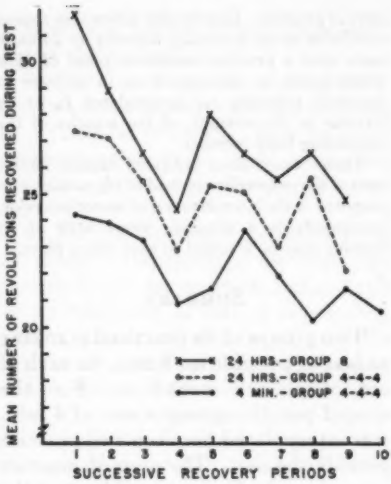


FIG. 3. Mean gain in rate of responding after rest as a function of successive recovery periods

8, and 4.53 for Group 4-4-4, $df = 9$ and 207, and 9 and 225, respectively).

The spacing of practice treatment has resulted in a slightly, but significantly, higher terminal level (Trial 24) of responding for the group with a 4-min. rest, for, as noted previously, the difference between Trial 24 means is significant on Day 1. The difference between the groups, however, is nowhere as great as on this day.

Since (*a*) response rate during the middle and terminal portions of the practice session tends to increase over successive days, and (*b*) level to which recovered as a function of 24-hr. rest decreases slightly, it follows that the amount of 24-hr. spontaneous recovery for the spaced- and massed-practice groups should decrease as a function of days.[3] The mean 24-hr. spontaneous recovery, for successive days is plotted in Fig. 3. The mean day-to-day recovery after 4 min. of rest for Group 4-4-4 is shown in the lower portion of Fig. 3.

Although the curves are quite irregular, the downward trends of all three functions are nevertheless apparent. Some of the irregularity of each of the three functions appears associated with the weekend of rest (72 hr.), for the weekend precedes or corresponds to the fifth recovery period for the two 24-hr. curves and the sixth for the 4-min. curve. The negative slope to the 4-min. recovery function follows from (*a*) above, since the level of recovery after 4-min. rest remains constant over days. The $F$ based upon the day-by-day Trial 13 means for between-day variance over day-by-subject variance failed to reach the .05 level of confidence with 9 and 225 *df* ($F = 1.12$).

## DISCUSSION

Hull (**4**) has assumed that the amount of reactive inhibition ($I_R$) is an increasing function of the number of prior responses per unit time. If the decrements in rate of cranking can be attributed to reactive inhibition, then a number of the present findings can be used to test some of the implications of $I_R$ theory.

Hull has further suggested that with the dissipation of $I_R$ the resting response (${}_sI_R$) is strengthened and that the resting response be given the properties of habit (${}_sH_R$). Some of these properties of $I_R$ and ${}_sI_R$ have been tested by Kimble (**5**) and in general the results have tended to verify many of the expectations concerning the two response-inferred variables.

[3] It should be understood that 24 hr. of rest were interpolated between successive days of practice except for one 72-hr. rest between the fifth and sixth days of practice. Thus, the reference *24-hr. rest* is not always appropriate.

In order to test or discover some of the properties of $I_R$ and ${}_sI_R$ by means of the present data it is necessary to assume that the cranking *habit* (${}_sH_R$ as distinguished from ${}_sE_R$) does not increase in strength as a function of number of

previous responses. It seems reasonable to make this assumption at present, yet it should be clear that it is made tentatively.

If the resting response is strengthened during rest as $I_R$ is dissipated, the initial rate of responding should decrease as a function of the number of days of practice. The results confirm this prediction, yet the magnitude of the decrease on the tenth day of practice on an extremely effortful task was surprisingly low. There are, of course, the possibilities that the decremental effect was almost completely overcome by an opposed increase in (*a*) the strength of the cranking habit through learning, or (*b*) response through increased strength of the musculature. The finding that the rate of responding during the middle and terminal portions of practice increased as a function of days suggests that (*b*) is probably correct. This can be interpreted as the growth of muscular strength quite apart from increasing response strength through learning. With this interpretation it is easier to understand why the difference in initial response rate on Day 10 versus Day 1 was not larger.

Two other findings of importance are that (*a*) the level of responding on Trial 1 on each day for the spaced group was remarkably similar to that for the massed group, and (*b*) the spaced group recovered to the same level of responding after four minutes of rest on each day of practice. Since Group 4–4–4 both responded more often and rested more often than the massed group, more opportunity was afforded $I_R$ to dissipate. Consequently, the spaced group might be expected to generate more ${}_SI_R$. Neither (*a*) nor (*b*) above, however, suggests that more ${}_SI_R$ has been accrued by the spaced-practice group. In contrasting these two groups the argument that augmented muscle strength and/or learning equaled a growing and permanent resting response (decrement) can still be advanced. This seems unlikely or at least quite unparsimonious for, if true, both the positive and negative reaction tendencies must grow at rates such that their magnitudes cancel when added and this must obtain on each of the ten days of practice.

The interpretation of the massing and spacing data is, of course, limited by the range of distribution chosen for study. Nevertheless, the range of inquiry has not been markedly restricted. It seems reasonable for the present to conclude that some distributions of practice phenomena are transitory and temporary in character. The results of this investigation have shown little differential effect ascribable to the manner of distribution of practice on preceding days. This, too, seems to be characteristic of results found with motor *learning* tasks (6).

The results have indicated systematic differences in performance as a function of number of days of practice. Day-by-day differences appear ascribable to an increasing capacity to do more work over a practice session—a point of view which might be interpreted as an increase in threshold tolerance for accumulated $I_R$ or an increase in the strength of the muscles of the responding body member.

These results have indicated further limitations of the properties ascribed to ${}_SI_R$ as currently proposed in the interpretation of motor behavior, particularly in a situation where little or no learning may be assumed to have taken place.

## Summary

Two groups of *S*s practiced cranking as fast as possible for 8 min. on each of ten successive weekdays. For the spaced-practice group a rest of 4 min. was interpolated between two practice periods of 4 min. The massed-practice group cranked continuously for the entire 8 min.

The data were presented as mean number of revolutions per successive 20-sec. scoring periods on each day of practice.

Continuous cranking led to marked decrements in rate of response during each experimental session. The major findings are listed below.

1. The initial rate of responding for each day decreased significantly but by a small amount as a function of days of practice. There appeared to be no difference as a function of spacing of prior practice.

2. Rate of responding at 4 or 8 min. was an increasing function of days of practice. The difference between the groups at 8 min. was attributable to the preceding 4-min. rest period.

3. The level to which the spaced-practice group recovered after 4 min. of rest remained invariant over days.

4. The amount of 24-hr. spontaneous recovery decreased as a function of days for both massed- and spaced-practice groups. The difference between the groups was related to the

difference in terminal rate of responding of the preceding day.

5. The 4-min. spontaneous recovery for the spaced-practice group decreased as a function of days. This decrease appears related to the increasing response rate immediately prior to rest.

6. For the extremes of distribution of practice represented by the two groups it can be concluded that little or no difference in rate of responding over a number of days can be attributed to the spacing of practice on preceding days. Differences in performance over a succession of days appear to be influenced more by the number of days of prior practice than by distribution of this practice.

(Manuscript received December 29, 1951)

## References

1. Bilodeau, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. San Antonio, Tex., Human Resources Research Center, Lackland Air Force Base, December 1951. (*Research Bulletin* 51–34.)
2. Bilodeau, E. A. Massing and spacing phenomena as a function of prolonged and extended practice. San Antonio, Tex., Human Resources Research Center, Lackland Air Force Base, February 1952. (*Research Bulletin* 52–9.)
3. Bilodeau, E. A. Performance decrement in a simple motor task before and after a single rest. *J. exp. Psychol.*, 1952, **44**, 381–390.
4. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
5. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
6. Reynolds, J. B. Private communication.

# A NEW METHOD FOR ANALYZING PRINTED ENGLISH [1]

EDWIN B. NEWMAN AND LOUIS J. GERSTMAN [2]

*Harvard University*

Language, as a form of human behavior, has long interested psychologists. Its scientific study has employed all of the methods which could be borrowed from other academic disciplines, as well as a fair assortment of the psychologist's own stock-in-trade. And yet language is so complex that even our most powerful methods of analysis do not take us far.

In many respects analysis at the level of words would be the most useful. Words come closest, at least, to the layman's idea of what are the natural units of speech and communication. But the difficulties are very great because the number of words is large and it is not easy to get homogeneous samples of sufficient size. It is understandable, therefore, that the first application of new tools should be made at the level of letters rather than words.

The present paper contains certain proposals for a new method which may prove to be both practical and fairly useful in estimating how much structure or pattern a sample of language possesses. For the reasons given above, the application of the method has been limited to sequences of letters in ordinary text. In addition to describing the method, the results of an analysis of printed English are reported since the argument for the method depends largely upon the demonstration that it works in practice.

[1] This study has been carried out with assistance from the Human Resources Research Laboratory under contract AF33(038)14343, Psychological Laboratory Report PLR-2.

[2] Now at Haskins Laboratory, New York, N. Y.

## Information in Printed English

The point of departure for this study is the estimation of the "information" contained in printed English by Shannon (6, 7, 8). The term "information" comes to have a quite specific and technical meaning. In general, a message such as a sequence of printed symbols contains information to the extent that it directs the receiver of the message to choose one of a specified set of alternatives, whether they be objects, or actions, or more complex meanings. Each of the letters of our printed alphabet might, if we made the best possible use of it, direct the reader to one of 26 equally probable alternatives, and a series of $n$ letters could designate $26^n$ alternatives.

In practice, we know that the use of letters is very far from optimal. For one thing, they were originally devised to represent the sounds of speech, not objects or ideas. For another, they are not used with equal frequency, and there are many combinations of letters which never occur in the normal use of language. Note, for example, the empty cells in Table 1 below. It becomes necessary to estimate how severe are these restrictions and over how long a sequence their effects are felt. The greater is the extent of these constraints, the less will be the information conveyed by a message of any given length.

The straightforward approach to the problem would require a tabulation of all of the sequences of letters which occur in the sample of language in which we are interested, and from this, an estimation of the probability which would attach to each such sequence. If we are content with single letters, with sequences of two or of three letters, this kind of tabulation can be, and has been, done. But the work required goes up in geometric proportions. There are roughly a hundred million million possible arrangements of 10-letter sequences, so that it is evidently impossible to extend this type of analysis to series of this length. Clearly, the direct, frontal approach must be given up.

Shannon has recently employed an interesting variant of an old game in order to gain some estimate of the information contained in longer series (7, p. 54). The device employs a kind of reversal, a dual, of the usual communication link. In place of transmitted information, the store of a subject's verbal habits is drawn upon. He is given a series of letters from any convenient text and is asked to guess what is the next letter. He is either corrected and goes on, or he continues

to guess until the actual letter is found. In either case, the extent to which the guessing is better than chance indicates the extent to which the subject is constrained by his built-in verbal habits. Were he able to guess perfectly, the new letter could convey no new information. If his guess were random, the new letter could bring maximum information. From a limited experiment conducted in this manner, Shannon has constructed a curve showing the upper bound on the information added by each additional letter in a chain constrained by the previous ($n$–1) letters.

There is an alternative procedure with which it is possible to test for dependencies over longer spans. This is the autocorrelation function which one of us has recently applied to sequences of vowels and consonants (4). By calculating the sum of the cross-products of all pairs of values in a time series which are separated by any given interval, a measure is obtained which expresses, in our present terminology, the dependence of the second value on the first. This is not the place to discuss the relative merits of these two basically different approaches, information and autocorrelation, for the problem of communication. Two differences only are important for our present problem. First, with finite series the autocorrelation function throws away certain information that is retained in the informational analysis. This is a result of disregarding the intermediate values in a series and taking the cross-products across an unspecified gap. On the other hand, this is not as serious as it first seems, for, as Elias has recently pointed out, the autocorrelation sets an upper bound on the information in a time series (1).

There is a second and more serious difficulty that cannot be surmounted so easily. This is the clearly nonmetric character of the letters of the alphabet. The autocorrelation function requires that the values in the time series can be placed on some kind of a scale, usually a full-fledged ratio scale. It is hard to conceive any way in which letters can be ordered such that their "cross-product" would be meaningful. The only thing we dan do is to map them into a kind of multidimensional probability space in which there are as many dimensions as letters. Such a model gains us nothing.

## The Coefficient of Constraint

In the absence of any practical statistic to use in analyzing sequences of letters, we have turned to an alternative measure which will be called the *coefficient of constraint*. The coefficient of constraint is a measure which behaves roughly in the same way as does the coefficient of correlation, that is, it has a value of zero when there is independence of the two measures, and a value of one when there is complete dependence.[3] It differs from the coefficient of correlation in that it may have only positive values. To give some intuitive notion of what is involved, we shall first describe the procedure which was followed.

The basic step consists of the construction of a set of contingency tables, one for each step of displacement, much in the same way that scatter diagrams might be constructed for the autocorrelation function. Each table has 28 × 28 cells (26 letters plus space and period). The first table shows the frequency with which each letter is followed immediately by each other letter. The next table contains the letters following a given letter skipping over one, as AxA, AxB, etc. The third table skips two letters, thus, AxxA, AxxB, and so on for the fourth and fifth tables.

The measure which is proposed describes the *additional* information in the second of the pair of letters compared with the information which that second letter might contain if it occurred alone. This is already a familiar computation, for it is identical with the way in which a measure of redundancy is found.

[3] As is nearly always the case, the measure which is proposed here bears a generic relationship to certain alternative expressions. Shannon and others have defined a "distance" in probability space which is the converse of our "constraint" (Cybernetics Symposium, AAAS meeting, New York, December 28, 1949). Cf. also, Parke and Lamson (5) who casually use the term "coefficient of correlation" much as we use "coefficient of constraint." It is probable that there are others. No attempt has been made to survey the literature exhaustively.

To put the matter in more formal terms, the information contained in the single letters, assuming that they are independent of one another, is given by:

$$H = -\sum_{1}^{a} p(i) \log_2 p(i), \qquad (1)$$

in which $H$ is the measure of information, expressed in bits, and $p(i)$ is the empirical probability associated with one of the $a$ alternatives. The information contained in single letters is easily computed by the application of this formula to the empirical probabilities in the sample under consideration.

Each of the contingency tables may be treated in the same way. If the two selections, that is, for rows and columns, have been made independently, the table will contain twice the information found for the single letters. The actual information is calculated as before:

$$H(1,2) = -\sum_{1}^{a}\sum_{1}^{a} p(i,j) \log_2 p(i,j). \qquad (2)$$

The expression, $H(1,2)$, is read to mean the information contained in pairs of adjacent items in a sequence. From this it follows that the information added by the second letter of the pair is measured by the difference between these two estimates, or:

$$H_1(2) = H(1,2) - H(1), \qquad (3)$$

which may be read, the information in item (2) when (1) is given is equal to the total information contained in the pairs (1,2) minus the information in (1) alone. The more general form of this expression would be:

$$H_1(n) = H(1,n) - H(1). \qquad (3a)$$

It is now possible to define the *coefficient of constraint*, $D$, as:

$$D(n) = 1 - \frac{H_1(n)}{H(1)}. \qquad (4)$$

In general, zero constraint will occur when the $n$th selection is independent of the first selection, or:

$$H_1(n) = H(1). \qquad (4a)$$

Similarly, the maximum value would be given if each column were uniquely determined by each row, in which case:

$$H_1(n) = 0, \qquad (4b)$$

and $D$ would become unity. It should also be noted in passing that $D(2)$ is identical with the usual measure of *relative redundancy* for the second letter in a sequence of two letters.

## Measured Constraint in Printed English

As a test of the usefulness of our method we have examined a 10,000-letter sample of printed English. The text used was taken from the Bible and consisted of the larger part of Isaiah xxix–xxxi in the King James version. The material was punched into Teletype tape using the 26 letters plus one symbol for space and a second symbol for period. No space was used following a period, and other punctuation was disregarded. The contingency tables were constructed by running the tape through our present Autocorrelator and Sequence Tabulator, a device that is a substantial elaboration of the device that has been previously described (**3**). However, even with the present machine, 28 runs were required for each table so that the total number of operations for the six tables here reported ran to about 1,500,000. This was more than enough to test the adequacy of our maintenance procedures!

Six contingency tables were constructed by these methods. They were for steps 1–2, 1–3, 1–4, 1–5, 1–6, and 1–10. Only two of the tables are presented here, Table 1 for the first

TABLE 1

FREQUENCY WITH WHICH EACH LETTER ACROSS TOP OF TABLE WAS FOLLOWED IMMEDIATELY BY EACH LETTER AT LEFT
(Total sample of 10,000 letters from King James Bible, Isaiah xxix–xxxi)

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | Spc. | Prd. | Tot. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | | 7 | 16 | 15 | 67 | 13 | 10 | 151 | 5 | 5 | 3 | 17 | 25 | 8 | 1 | 9 | | 22 | 20 | 24 | 1 | 3 | 16 | 1 | 2 | | 231 | 20 | 692 |
| B | 6 | | | | 3 | | | | 3 | | | | 1 | | 6 | | | | | | | | | | | | 86 | 11 | 116 |
| C | 19 | | | | 16 | | | | 12 | | | | | 6 | 4 | | | 3 | 7 | 6 | 2 | | | 1 | | | 56 | 1 | 133 |
| D | 19 | | | 4 | 45 | | | | 14 | | | 17 | | 153 | 8 | | | 32 | 1 | 1 | 11 | | | | | | 65 | 2 | 372 |
| E | 4 | 54 | 24 | 50 | 31 | 6 | 13 | 294 | 18 | 2 | 27 | 48 | 57 | 49 | 5 | 25 | | 118 | 38 | 34 | 2 | 41 | 14 | | 21 | | 29 | 1 | 1005 |
| F | 3 | | | | 14 | 6 | | | 3 | | | 4 | | 1 | 80 | | | 5 | | 2 | | | | | | | 67 | 17 | 202 |
| G | 12 | | | 5 | 3 | | 1 | | 10 | | | | | 47 | 1 | | | 3 | | | 18 | | | | | | 20 | 1 | 121 |
| H | 1 | | 26 | | 7 | | 25 | | | | | | | | | 4 | | | 81 | 379 | | | 37 | | | | 127 | 4 | 691 |
| I | 24 | 1 | 5 | 13 | 12 | 12 | 2 | 70 | | | 2 | 27 | 13 | 12 | 5 | 5 | | 40 | 11 | 17 | 9 | 9 | 53 | | 3 | 3 | 110 | 9 | 467 |
| J | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | 12 | 1 | 14 |
| K | 19 | | 3 | | 2 | | | | 2 | | | 2 | | 8 | 7 | | | 11 | | | | | | | | | 5 | | 59 |
| L | 99 | 7 | 4 | | 32 | 8 | 3 | | 26 | | | 103 | | 2 | 12 | 13 | | 1 | 5 | 7 | 18 | | 1 | | | | 60 | 1 | 402 |
| M | 16 | | | | 29 | | 4 | | 19 | | | | 6 | | 27 | | | 5 | 5 | | 11 | | 1 | | | | 66 | 1 | 190 |
| N | 161 | | | | 58 | | | | 118 | | 3 | | | 3 | 67 | | | 19 | 3 | | 37 | | 10 | | | | 40 | 1 | 520 |
| O | | 6 | 33 | 17 | 12 | 44 | 8 | 34 | 17 | 2 | | 40 | 15 | 37 | 10 | 18 | | 49 | 14 | 52 | | 5 | 28 | | 10 | | 115 | 4 | 570 |
| P | 3 | | | | 19 | | | | 4 | | | | 6 | | 9 | 2 | | 2 | 10 | | 18 | | | | 2 | | 52 | | 127 |
| Q | | | | | | | | 1 | 2 | | | | | | | | | | | | | | | | | | | | 3 |
| R | 63 | 15 | 9 | 17 | 100 | 16 | 10 | 12 | 17 | | | | | | 102 | 20 | | 8 | 4 | 25 | 28 | | 2 | | 1 | | 27 | | 476 |
| S | 39 | 1 | | 9 | 40 | | 1 | 1 | 80 | | 2 | 6 | 2 | 19 | 8 | 3 | | 20 | 16 | 8 | 24 | | | | 1 | | 169 | 3 | 452 |
| T | 92 | | 7 | | 40 | 1 | 2 | 16 | 97 | | | 11 | | 30 | 32 | 13 | | 13 | 77 | 11 | 29 | | | | | | 326 | 27 | 824 |
| U | 9 | 12 | 5 | 4 | | 8 | 1 | 5 | | 5 | | 1 | 9 | 1 | 75 | 2 | 3 | 18 | 9 | 10 | | | | | | | 37 | | 214 |
| V | 12 | | | | 11 | | | | 4 | | | 2 | | | 15 | | | 2 | | | | | | | | | 12 | | 58 |
| W | 7 | | | 1 | 2 | | | | | | 1 | | | | 29 | | | | 3 | | | | | | | | 117 | 12 | 172 |
| X | 1 | | | | 2 | | | | | | | | | | | | | | | | | | | | | | | | 3 |
| Y | 33 | 8 | 1 | 1 | 21 | 2 | 2 | 4 | | | | 6 | 4 | 1 | 2 | | | 8 | 1 | 13 | | | | | | | 26 | 4 | 137 |
| Z | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | | 3 |
| Space | 48 | 3 | | 225 | 414 | 83 | 34 | 99 | 16 | | 19 | 115 | 47 | 134 | 65 | 11 | | 92 | 139 | 209 | 5 | | 8 | 1 | 90 | | | | 1857 |
| Period | 1 | 2 | | 11 | 24 | 3 | 5 | 4 | | | 2 | 3 | 5 | 9 | | 2 | | 5 | 8 | 26 | 1 | | 2 | | 7 | | | | 120 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 10,000 |

TABLE 2

FREQUENCY WITH WHICH EACH LETTER ACROSS TOP OF TABLE WAS FOLLOWED AFTER NINE LETTERS BY EACH LETTER AT LEFT

(Total sample of 10,000 letters from King James Bible, Isaiah xxix–xxxi)

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | Spc. | Prd. | Tot. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 63 | 8 | 8 | 23 | 69 | 10 | 8 | 42 | 42 | 3 | 3 | 29 | 15 | 30 | 48 | 10 | | 27 | 34 | 49 | 13 | 3 | 12 | | 10 | 2 | 122 | 8 | 691 |
| B | 5 | | 4 | 3 | 21 | 2 | 3 | 3 | 6 | | 1 | 4 | 2 | 3 | 8 | 1 | | 6 | 7 | 12 | 3 | | 2 | | 2 | | 14 | 2 | 116 |
| C | 10 | 2 | 1 | 10 | 13 | 3 | 1 | 13 | 6 | | | 8 | 2 | 9 | 8 | 2 | | 7 | 8 | 11 | 9 | 2 | 4 | | 5 | | | 2 | 134 |
| D | 28 | 7 | 6 | 7 | 44 | 12 | 8 | 23 | 15 | | 3 | 17 | 6 | 19 | 16 | 7 | | 18 | 21 | 22 | 6 | 2 | 5 | | 9 | | 67 | 4 | 372 |
| E | 58 | 15 | 18 | 38 | 110 | 17 | 8 | 62 | 58 | 1 | 8 | 44 | 24 | 62 | 61 | 7 | | 41 | 33 | 80 | 25 | 3 | 20 | | 20 | | 181 | 12 | 1006 |
| F | 19 | 7 | 3 | 3 | 24 | 4 | 1 | 21 | 10 | 1 | 2 | 15 | 6 | 9 | 17 | 2 | | 5 | 4 | 12 | 1 | 3 | 1 | | | | 30 | 2 | 202 |
| G | 9 | 1 | 2 | 1 | 16 | 2 | 1 | 8 | 6 | | | 6 | 2 | 6 | 5 | 1 | | 4 | 7 | 15 | 4 | 3 | 1 | | 2 | | 16 | 3 | 121 |
| H | 46 | 3 | 11 | 24 | 72 | 11 | 10 | 56 | 18 | | 6 | 29 | 11 | 39 | 37 | 12 | | 29 | 27 | 63 | 14 | 6 | 16 | | 7 | | 127 | 17 | 691 |
| I | 40 | 7 | 2 | 16 | 44 | 12 | 6 | 35 | 22 | | 3 | 25 | 10 | 29 | 27 | 3 | | 25 | 25 | 27 | 8 | 4 | 9 | 1 | 3 | | 78 | 5 | 466 |
| J | | | 1 | | | | | 1 | | | 1 | | 1 | | 1 | 1 | 1 | | 1 | 1 | 2 | | | | 1 | | 2 | | 14 |
| K | 2 | | 1 | 6 | 4 | 2 | | 4 | 2 | | | 6 | | 3 | 2 | | | 6 | 1 | 3 | 3 | | | | | | 14 | | 59 |
| L | 27 | 4 | 3 | 18 | 40 | 8 | 4 | 28 | 25 | 1 | 1 | 17 | 9 | 24 | 19 | 8 | | 13 | 14 | 38 | 11 | 2 | 7 | 1 | 6 | | 67 | 8 | 403 |
| M | 9 | 3 | 3 | 8 | 15 | 5 | 1 | 12 | 3 | | | 4 | 5 | 11 | 14 | 2 | | 17 | 8 | 16 | 5 | 2 | 4 | | 2 | | 39 | 2 | 190 |
| N | 30 | 8 | 11 | 14 | 47 | 16 | 6 | 42 | 26 | | 3 | 15 | 9 | 25 | 37 | 5 | | 26 | 19 | 42 | 7 | 1 | 6 | | 3 | | 117 | 5 | 520 |
| O | 45 | 7 | 5 | 20 | 64 | 12 | 8 | 49 | 27 | 1 | 1 | 33 | 8 | 16 | 27 | 11 | | 36 | 18 | 44 | 14 | 5 | 6 | | 8 | | 100 | 5 | 570 |
| P | 12 | 1 | 2 | 11 | 11 | 2 | 2 | 5 | 3 | | | 5 | 5 | 6 | 4 | 9 | | 7 | 2 | 6 | 3 | | 1 | | 2 | | 27 | 1 | 127 |
| Q | | | | | | | | 1 | | | | 1 | | | | | | | | | | | | | | | | | 3 |
| R | 27 | 3 | 7 | 27 | 49 | 8 | 8 | 27 | 16 | | 2 | 19 | 6 | 31 | 33 | 6 | | 20 | 29 | 35 | 7 | 4 | 14 | | 7 | | 89 | 4 | 478 |
| S | 30 | 2 | 7 | 17 | 39 | 7 | 13 | 40 | 17 | | 2 | 13 | 10 | 26 | 25 | 1 | | 18 | 18 | 46 | 7 | | 8 | | 8 | | 94 | 4 | 452 |
| T | 58 | 10 | 9 | 28 | 71 | 20 | 7 | 62 | 41 | 3 | 3 | 24 | 16 | 46 | 55 | 8 | | 43 | 47 | 85 | 17 | 6 | 13 | 1 | 6 | | 137 | 8 | 824 |
| U | 16 | 1 | 3 | 8 | 19 | 8 | 2 | 11 | 15 | 1 | | 7 | 5 | 11 | 9 | 4 | | 11 | 11 | 20 | 3 | 1 | 3 | | 3 | | 40 | 1 | 213 |
| V | 5 | | | 2 | | 2 | | 7 | 3 | | 1 | 2 | 2 | 1 | 1 | 1 | | 3 | 5 | 4 | 1 | 1 | 1 | | 3 | | 12 | 2 | 59 |
| W | 14 | 4 | 4 | 6 | 20 | 4 | 3 | 11 | 3 | | 3 | 9 | 2 | 7 | 12 | 2 | | 11 | 8 | 19 | 4 | 1 | 2 | | 2 | | 16 | 5 | 172 |
| X | | | 1 | | | | | | | | | 1 | | | | | | | 1 | | | | | | | | | | 3 |
| Y | 4 | 1 | 1 | 9 | 16 | 5 | 3 | 13 | 4 | | 1 | 6 | 1 | 15 | 7 | 1 | | 5 | 10 | 7 | 3 | 1 | 4 | | 1 | | 18 | 1 | 137 |
| Z | | | | | | | | | 1 | | | | | | 1 | | | | | | | | | | | 1 | | | 3 |
| Space | 129 | 21 | 21 | 70 | 186 | 29 | 18 | 111 | 86 | 1 | 15 | 63 | 31 | 82 | 83 | 21 | 2 | 94 | 89 | 160 | 40 | 8 | 32 | | 25 | 1 | 418 | 18 | 1854 |
| Period | 5 | 1 | | 3 | 12 | 1 | | 4 | 10 | 2 | | 1 | 2 | 10 | 13 | 2 | | 6 | 4 | 7 | 3 | 1 | 1 | | 2 | | 29 | 1 | 120 |

step, 1–2, and Table 2 for the last one, 1–10. Table 1 possesses some interest in itself for it gives in detail a tabulation of the digraphs in this sample. Since space and period are included, the table also shows, on a restricted sample, what is the distribution of letters which begin words and end words, and begin sentences and end sentences. The counts for words are obtained by combining space and period, of course. Moreover, there are 1977 spaces and periods, and hence 1978 words in the sample, and 120 sentences. This gives an average word length (not counting the space) of just over four letters, and a sentence length of 16.5 words.

The results of the entire calculation are summarized in Table 3. It gives in the first column $H(1)$ (the same for all steps), in the second column the dependent "information," $H_1(n)$, and in the third column the computed coefficient of constraint, $D(n)$.

A rather striking regularity emerges when the values of $D$ are plotted on logarithmic scales as in Fig. 1. In this plot the abscissa is the step that is compared with step (1). The first point must represent a value of $D$ equal to unity by definition. The first calculated point is plotted above the value (2), and so on for the other five values. The straight line which has been drawn has a slope of $-2.0$.

TABLE 3

COEFFICIENT OF CONSTRAINT ($D$) FOR SIX STEPS IN THE SEQUENTIAL ANALYSIS OF PRINTED ENGLISH TEXT

| Step | $H(1)$ | $H_1(n)$ | $D(n)$ |
|---|---|---|---|
| 1–2 | 4.08 bits | 3.17 bits | .223 |
| 1–3 | 4.08 | 3.66 | .103 |
| 1–4 | 4.08 | 3.82 | .064 |
| 1–5 | 4.08 | 3.92 | .039 |
| 1–6 | 4.08 | 3.97 | .027 |
| 1–10 | 4.08 | 4.03 | .012 |

It is obvious that the straight line is an excellent approximation to the entire set of observed points and, with the possible exception of the second point, there is no significant deviation.

If it is assumed that this straight line adequately describes the results, it is possible to sum them up in even more simple fashion by writing the empirical equation:

$$D(n) = 1/n^2. \qquad (5)$$

As we shall show later, the rather simple form of this equation leads to an important simplification of the estimated information in a long sequence of symbols.

## A New Estimate of the Upper Bound of Information in Printed English

With these empirical results in hand, we are emboldened to propose a new estimate of the amount of information in continuous text. In the actual use of language, the constraints which operate invariably extend over sequences much longer than one or two letters. What we are interested in discovering is the information contained in each letter when the series is, for all practical purposes, infinitely long. Presumably, as the transitional probability depends on a longer and longer sequence, the information added by the final letter in the sequence approaches the average information over the entire sequence. Using the symbol $F(n)$ for the information in the last letter of a sequence of $(n)$ letters, it can be shown that:

$$F(n) = H_{i,\,j,\,k\,\ldots\,n-1}(n) \rightarrow H(i, j, k \ldots n)/n. \qquad (6)$$

Since it is quite impractical to evaluate either the successive values of $F(n)$ or the limiting value represented by the right-hand side of the expres-

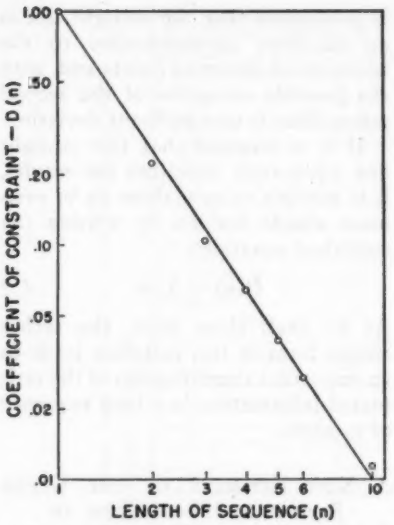


FIG. 1. The constraint exercised by one letter in printed English on a second letter following it after a varying interval. The interval is measured by the number of letters in the sequence, the first or constraining letter being number one. Both $n$ and $D(n)$, the coefficient of constraint, are plotted on logarithmic scales.

sion, some alternative must be found. Shannon has worked out one approach by deriving both an upper and a lower bound from the results of the guessing game. The reader is referred to Shannon's paper for the details of the method (7).

As a second alternative, we are proposing that the coefficient of constraint be regarded as a proportion by which the amount of information per letter is successively reduced as the constraining sequence is made longer and longer. Using the term $F(n)$ as defined in equation (6) above, that is, as the additional information added by the $n$th letter in a sequence of $n$ letters, we would write:

$$F(n) \leqq (1-D(n))\cdot F(n-1). \quad (7)$$

It is possible to show that this relation holds in several limiting cases. For one thing, the equality holds in the case of the first step in the sequence. Since $D(2)$ is the relative redundancy and $F(1)$ is the information in the first letter of a sequence, it follows as a matter of definition that:

$$F(2) = (1-D(2))\cdot F(1). \quad (7a)$$

Furthermore, we may define the behavior of both measures if it is assumed that any item is independent of the $n$th preceding item. In this case we have the known limiting equality

$$F(n) = F(n-1). \quad (8)$$

At the same time, we have by definition zero constraint so that equation (7) reduces to equation (8). Finally, if $D(n)$ should be unity, equation (7) says that $F(n)$ becomes zero. This outcome is intuitively correct and we omit the somewhat tedious demonstration that $F(n)$ should indeed behave this way. In general, however, no proof of equation (7) has been found and there are limiting cases in which it is proved not to apply.

If it is granted for the moment that the argument is at least intuitively correct, there are then two methods of procedure open to us. The first would

TABLE 4

COMPARISON OF VARIOUS ESTIMATES OF THE ADDITIONAL INFORMATION, $F(n)$, ADDED BY THE $n$TH LETTER IN A SEQUENCE

| Step | Shannon Calculated | Shannon Smoothed Curve | Present Paper |
|---|---|---|---|
| 1 | 4.03 | 4.05 | 4.08 |
| 2 | 3.42 | 3.45 | 3.17 |
| 3 | 3.0 | 3.0 | 2.84 |
| 4 | 2.6 | 2.7 | 2.66 |
| 5 | 2.7 | 2.5 | 2.56 |
| 6 | 2.2 | 2.3 | 2.49 |
| 7 | 2.8 | 2.2 | 2.44* |
| 8 | 1.8 | 2.1 | 2.40* |
| 9 | 1.9 | 2.0 | 2.37* |
| 10 | 2.1 | 2.0 | 2.34 |
| 100 | 1.3 | 1.9(?) | 2.06 |

* By interpolation from Fig. 1.

be the direct computation of the successive values of $F(n)$. These are given in Table 4 together with a comparison of the values read from Shannon's smoothed curve, and his observed values. It is reasonably clear that the agreement is fairly good although it is not possible to give any adequate estimate of how good since there is no alternative view against which to test it.

The second possibility arises from the fact that it was possible to derive an empirical equation (5) for $D(n)$ above. Let us note first that equation (7) may be expanded as follows:

$$F(n) \leqq F(1)\cdot(1-D(2)) \quad (1-D(3)) \cdots (1-D(n)). \quad (9)$$

Substituting, according to equation (5) and replacing $F(1)$ with $H(1)$, we get:

$$F(n) \leqq H(1)\ (1-1/2^2) \cdot(1-1/3^2) \cdots (1-1/n^2). \quad (10)$$

By rearranging terms, this becomes, successively,

$$F(n) \leqq H(1)\ [(2^2-1)/2^2] \cdot[(3^2-1)/3^2] \cdots [(n^2-1)/n^2] \quad (10a)$$

$$\leqq H(1)\ [3\cdot4\cdot5 \cdots (n+1)] \quad [1\cdot2\cdot3 \cdots (n-1)]/ \quad [2^2\cdot3^2\cdot4^2 \cdots n^2] \quad (10b)$$

$$\leqq H(1)\ (n+1)/2n. \quad (11)$$

An examination of this equation shows that it has several interesting properties. In the first place, it has a real limit, for as $n$ approaches infinity, the right side of the equation (11) approaches $H(1)/2$. This limit is exactly the value which we wanted above in equation (6). To be sure we believe that we have here only an upper bound to such a limit, but the available evidence suggests that the actual values do not fall too far below such an upper bound.

A second property of the equation is the fact that it rises to plus infinity for $n = 0$. It will be recalled that it has been customary to calculate a value for $H(0)$ by assuming that each of the 26 letters was represented with equal probability. Actually, a little reflection will show how arbitrary has been this assumption. It is just as possible to assume that the sender has a very large set of letters from which he might have drawn, and for which we can assume with equal justice an equivalent share of the total probability. The only thing we know is that he uses only 26 of these letters with a finite probability at the first (empirical) step. An extension of this reasoning leads rather easily to the acceptance of the notion of infinite information at the zero step, if indeed any meaning at all is to be attached to values of $n$ less than 1. The behavior of the function is consistent with this line of reasoning.

Finally, it becomes interesting to see whether this equation is particular to the form which equation (5) happened to take. A more general form of this equation might be:

$$D(n) = 1/n^b, \quad (12)$$

in which $b$ is an arbitrary slope constant from a plot such as Fig. 1. Except for the integral values of $b = 1$ and $b = 2$, there appears to be no general solution for the form of the function describing $F(n)$. Much more empirical evidence will be required before any great generality can be claimed for either equation (11) or equation (12).

The best summary of the results of this development is the plot in Fig. 2. In this are drawn Shannon's curve for the upper bound (dashed line) and our function according to equation (11) (solid line). Further, there are plotted the empirical values taken from Shan-

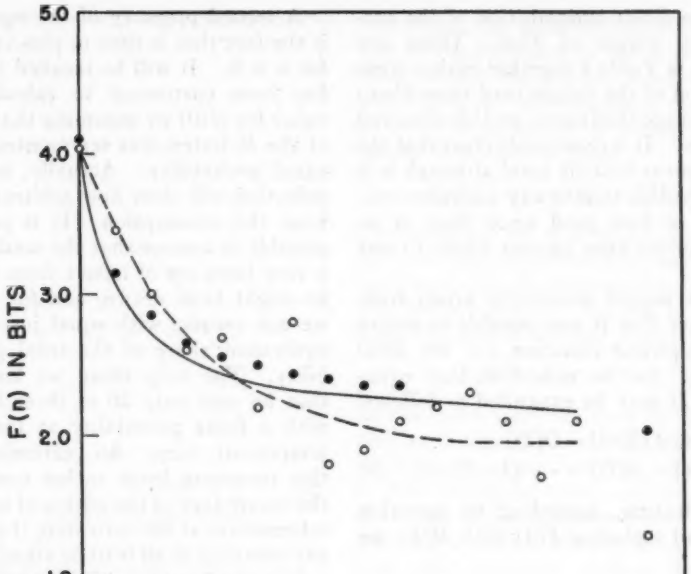


FIG. 2. The information added by the final letter in a sequence $n$ letters long. The figure shows in comparison both the experimental points (open circles) and the smoothed curve (dashed line) from Shannon (7) and the calculated points (solid circles) and computed function (continuous line) from the present paper. Our function has an asymptote at 2.04 bits.

non's article (open circles) and our results when calculated by the expanded form of equation (9) (solid circles). Within the range for which there are empirical data, there is no clear choice between the two functions. Extrapolation should be made with caution.

## CONFIDENCE LIMITS OF INFORMATION ESTIMATES

It is not unnatural that the psychologist should be particularly cautious about the confidence limits which attach to any extended observations of this sort. He has been trained always to inquire about the magnitude of sampling errors. Unfortunately, we know of no general solution to this problem for information measures, so that it seemed worth while to resort to somewhat more primitive methods. The obvious method of attack was by way of $\chi^2$ since the basic tables are effectively contingency tables. The following results are reported for what preliminary assistance they may be.

As a first step we computed both

$D(n)$ and $\chi^2$ for six sets of reduced tables formed by combining rows and combining columns of the large tables. Thus, we could form a $3 \times 3$ table in which the classification was Vowel-Consonant-Space. These reduced tables were $2 \times 2$, $3 \times 3$, and $4 \times 4$ in size. In every case, we found for a given sample and a given set of marginal totals that there was an almost precise linear relation between the two measures. There are several parameters of this relation at which we can guess roughly. One is the total $N$ of the sample. $D(n)$ is entirely independent of $N$ (except for sampling variation) whereas $\chi^2$ is proportional to $N$. Furthermore, $\chi^2$ is apparently proportional to $H(1)$ (or to log $a$, where $a$ is the number of categories).

Encouraged by this evidence, we have worked out the $\chi^2$ for a $10 \times 10$ table derived from Table 2 above. The reduction was accomplished principally by combining some of the less frequently used letters. For this table, $\chi^2$ equals 121 which gives a $p = .004$. Thus this $10 \times 10$ table shows a significant constraint, but by extrapolating it can be estimated that the full table for the 1 to 10 step probably does not differ significantly from chance. The full table for Step 6 would, on the other hand, certainly prove highly significant. Roughly it can be said that there are probably significant dependencies contained within the table for Step 10, but that a judicious and, one hopes, legitimate combination of columns is necessary in order to bring them out.

## Information and Linguistic Classification

It is not possible to make a study such as this without becoming curious about the relation of this estimate of the structure of printed language to other studies of lexical structure. Some of these relations are worth brief comment here.

One expectation which we had in making the study was not fulfilled. This was that there would appear a rather sharp drop in the coefficient of constraint about the fifth or sixth letter. This was based on the premise that the sequential dependencies connecting words would be ever so much weaker than the dependencies within words, and that the fifth and sixth letters would represent the beginning of new words. On the average, this appears not to be true. Perhaps the reason is that the break is blurred by various sorts of statistical variability. Words are not all the same length. There are some very high sequential dependencies among some words. Other explanations will occur to the reader. But the basic fact remains that there is very little evidence here that the order of magnitude of this dependency is different for words than for letters. If it were, we would have every reason to expect to find some sharp changes in slope somewhere about the length of the median word.

A second line of evidence seems to confirm this view. If our data are used to compute the amount of information *per word*, an estimate of 10.2 bits per word is obtained. This is simply the average word length, 5 letters, multiplied by the limiting value of information per letter, 2.04 bits. This would predict a vocabulary of about 1300 equally used words. But we know with some certainty that words are not used equally so that the vocabulary would have to be much larger. Actually, both Shannon (**7**) and Miller (**2**, p. 798) have independently taken the well-known harmonic function of Zipf (**9**) and calculated the presumed information per word from

the statistics of word use alone.[4] This gives estimates between 9.7 and 11.8 bits per word. Since we have not made any word counts on the material used here, it would appear that our estimate, based on letters, is as close as casual sampling would predict.

One final comparison has been made. By taking the sequential probabilities represented in Table 2, it is possible to compare the constraint exercised by letters within a word with the constraint on the initial letter of each word. This calculation shows that $H$ for initial letters in our sample is 4.00 bits, only slightly less than the 4.08 for all letters and spaces taken singly, and almost identical with the value of 4.01 calculated for single letters alone in this sample. Put in other terms, although the distribution of initial letters is quite different from distribution of all letters, knowledge of this distribution places no greater constraint on the choice of a letter than does the choice of a letter at random from the total sample. This is in sharp contrast to the constraint implied by the remainder of Table 1. With the initial letters removed, the $H$ is reduced from 3.17 to 2.96 bits per letter. It should be emphasized that in this paragraph, the constraint being discussed depends upon the *single* preceding letter. Our tables do not permit further analysis.

The consequence of this last finding must be taken together with the fact that the space itself is largely redundant. We know of no numerical results to demonstrate this, but the most casual inspection of guessing patterns will show that this is true. What is evidently happening, then, is that the space itself contributes little information, but the space plus the initial letter, when taken together, supply almost as much information as the average pair of letters.

In conclusion it can only be said that the present tools when applied to letters are not adequate to sense the somewhat subtle sequential dependencies among words. Actually the striking fact is that the constraints on letters are so much more marked than those among words. It seems probable that questions of grammatical structure and the like must be tackled directly.

[4] There seems to be some disagreement about the exact numerical value to be expected. Shannon says that the sum of the first 8727 ranks gives a total probability equal to unity. Miller, presumably using the definite integral, comes out with 22,000. Our own calculation, which takes into account the discontinuous values for the first 100 ranks, gives 12,370. On the other hand, Shannon gets a figure of 11.82 bits per word, Miller gives 10.6 bits per word, and our smaller $n$ only gets this estimate up to 9.7 bits per word. The reason for our disagreement with Shannon is not clear.

## Summary

1. A new measure, the coefficient of constraint, is proposed which provides a limited but extremely useful estimate of the sequential dependencies in sequences of events, such as letters, which cannot be ordered into a metric scale.

2. When the coefficient of constraint was calculated for a passage of printed English, a quite regular behavior of the coefficient was discovered which could be described by the expression $D(n) = 1/n^2$.

3. A possible use of the coefficient of constraint is proposed by which an upper bound is set on the average information per letter in the text examined. This upper bound has both considerable resemblance to, and a few striking differences from, Shannon's estimate of the same function.

4. Neither greater freedom nor greater constraint is discovered when

sequences of letters are examined of a length greater than that of one word.

(Manuscript received for priority publication March 13, 1952)

## References

1. Elias, P. A note on autocorrelation and entropy. *Proc. Inst. Radio Engrs*, 1951, **39**, 839.
2. Miller, G. A. Speech and language. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1951. pp. 789–810.
3. Newman, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.*, 1951, **64**, 252–262.
4. Newman, E. B. The pattern of vowels and consonants in various languages. *Amer. J. Psychol.*, 1951, **64**, 369–379.
5. Parke, N. G., III, & Lamson, E. W. *Distance and equivalence in sequence space*. Air Force Cambridge Research Laboratories, Report No. E4080.
6. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423; 623–656.
7. Shannon, C. E. Prediction and entropy of printed English. *Bell Syst. Tech. J.*, 1951, **30**, 50–64.
8. Shannon, C. E., & Weaver, W. *The mathematical theory of communication*. Urbana: Univer. of Illinois Press, 1949.
9. Zipf, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

# EXPERIMENTS ON SENSORY-TONIC FIELD THEORY OF PERCEPTION: V. EFFECT OF BODY STATUS ON THE KINESTHETIC PERCEPTION OF VERTICALITY [1]

SEYMOUR WAPNER AND HEINZ WERNER

*Clark University*

The previous papers of this series (**1, 2, 4, 6**) were concerned with organismic effects on the visual perception of verticality. It seems important to extend the experiments beyond visual perception for the following reason. It is necessary to inquire into the generality of organismic effects on objects independent of the modality in which they are perceived (**3**). Furthermore, the question arises whether specific peripheral factors, such as retinal displacement due to eye movements, etc., are capable of accounting for the effects on visual perception previously reported; or whether general organismic factors, irrespective of the involvement of peripheral factors, play a significant role. Consequently, experiments were carried out on the effect of "extraneous stimulation," i.e., stimulation from a source other than the object, on the perception of verticality as experienced kinesthetically.

In the present paper the results of three experiments on the kinesthetic perception of verticality are reported: (*a*) effect of head tilt; (*b*) effect of body tilt; (*c*) effect of concentric body rotation.

## Apparatus and Procedure

In all three experiments, the task for *S*, while *blindfolded*, was to adjust a rod to the vertical kinesthetically. The rod was 31 in. long; its lower end was pivoted at *S*'s waist level so that it rotated in the fronto-parallel plane. The angle of inclination to which *S* adjusted the rod was directly read from a protractor. The *S* grasped the rod with thumb and finger tips of both hands; he then had to move both hands simultaneously up and down the rod, and to change its angle of tilt until it appeared vertical to him.

*Experiment I: Sidewise head tilt.*—In this experiment the task was carried out under three conditions of head position: (*a*) sidewise tilt of head toward the left shoulder at an angle of 45°; (*b*) head upright; (*c*) sidewise tilt of head toward the right shoulder at an angle of 45°. For each of these main conditions two trials were given with the rod initially set in each of three starting positions: 30° Left; 0° (plumb line); and 30° Right. Thus a total of 18 trials was given to each *S*. Twelve men and 12 women served as *Ss*. The presentation of trials was randomized, and the data were analyzed as a three variable (head position, starting position, *Ss*) factorial design separately for men and women.

*Experiment II: Sidewise body tilt.*—In this experiment the task was carried out under three conditions of body position: (*a*) body tilted 30° L; (*b*) body upright; (*c*) body tilted 30° R. For the conditions of body tilt *S* rested his body, shoulder, and head against a leaning board of inverted **V** shape, cut out at the shoulders to permit free movement of the arms. Here, as in the first experiment, for each main condition three starting positions were used: 30° L; 0° (plumb line); and 30° R. Two trials were given for each combination of conditions making a total of 18 trials. The presentation was in random order. Twenty-one men served as *Ss*. The data were analyzed as a three variable (body tilt, starting position, *Ss*) factorial design.

*Experiment III: Body rotation.*—In this experiment the task was carried out under conditions of concentric body rotation. Labyrinthian stimulation was provided by a rotating chair described previously (**2**). Rotation was controlled by a constant speed motor through a Graham variable speed drive. One speed, 25 rpm, was used for both clockwise and counterclockwise motion.

The period of acceleration, from zero to a constant speed of 25 rpm, lasted 15 sec.; deceleration, from constant speed of 25 rpm to 0 rpm,

[1] This investigation was supported by a grant from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

TABLE 1

MEAN POSITION OF APPARENT VERTICAL*

| Sex | Head Position | | |
|---|---|---|---|
| | Left Tilt | Upright | Right Tilt |
| Men | +7.7° | −1.0° | −10.8° |
| Women | +6.4° | −1.4° | −9.3° |

* The plus sign indicates the position of apparent vertical clockwise of the plumb line (as viewed by *S*); the minus sign indicates the position of apparent vertical counterclockwise of the plumb line.

lasted 9 sec. The head was tipped forward so that a line from the corner of the eye to the ear canal was horizontal; it was fixed in position by means of a headrest. The rod was mounted on a horizontal board which, in turn, was fixed to the armrests of the rotating chair.

The kinesthetic adjustment of the rod to the vertical was carried out under five test conditions: (*a*) Control: absence of rotation (C); (*b*) Acceleration in a clockwise direction (A–CW); (*c*) Deceleration from constant clockwise rotation (D–CW); (*d*) Acceleration in a counterclockwise direction (A–CCW); (*e*) Deceleration from constant counterclockwise rotation (D–CCW).

The sequence of test conditions was: C; A–CW; D–CW; A–CCW; D–CCW. Between each of these conditions a 3-min. rest period was interposed. The sequence was repeated making a total of ten trials per *S*. In contradistinction to the two previous experiments only one starting position was used: the rod was set at objective vertical at the beginning of each trial. Six male *S*s were employed. The data were analyzed by a two-way classification (conditions, *S*s) analysis of variance.

*Measures employed.*—In all three experiments the position in which the rod was kinesthetically perceived as vertical (apparent vertical) was measured in degrees of deviation from the objective vertical (plumb line). Arbitrarily, an angular position—as viewed by *S*—clockwise of objective vertical is designated as "+" (tilted right), and counterclockwise as "−" (tilted left). The measure used to represent *S*'s performance is the algebraic mean of the adjustments for each condition.

## RESULTS

*Experiment I: Sidewise head tilt.*—The results of the effect of head tilt are summarized in Tables 1 and 2. For both men and women, the *F* test for the main experimental variable, head tilt, is significant beyond the .01 level of confidence. The mean position of the apparent vertical is rotated to the left of control (head upright) when the head is tilted right, and to the right of control when the head is tilted left. The means of apparent vertical under the three test conditions differ significantly from one another ($p < .01$) for both men and women.

*Experiment II: Sidewise body tilt.*—In general, the effect of body tilt is similar to that of head tilt, i.e., when

TABLE 2

ANALYSIS OF VARIANCE OF KINESTHETIC ADJUSTMENT OF ROD TO VERTICALITY UNDER HEAD TILT (EXP. I)

| Source of Variance | *df* | Women | | | Men | | |
|---|---|---|---|---|---|---|---|
| | | Sum of Squares | *F* | *p* | Sum of Squares | *F* | *p* |
| Head position (H) | 2 | 4449.40 | 74.21[a] | <.01 | 6185.36 | 33.34[c] | <.01 |
| Starting position (S) | 2 | 206.79 | 3.60[b] | >.05 | 21.47 | 1.49[d] | >.05 |
| Individuals (I) | 11 | 663.63 | 2.00[c] | >.05 | 532.30 | <1.00[c] | >.05 |
| H × S | 4 | 115.05 | 3.80[d] | <.01 | 39.53 | 1.37[d] | >.05 |
| H × I | 22 | 664.37 | 3.99[d] | <.01 | 2040.86 | 12.88[d] | <.01 |
| S × I | 22 | 212.32 | 1.28[d] | >.05 | 82.75 | <1.00[d] | >.05 |
| H × S × I | 44 | 332.51 | | | 316.92 | | |
| Total | 107 | 6644.07 | | | 9219.19 | | |

[a] Tested against pooled error (H × I and H × S).
[b] Tested against H × S.
[c] Tested against H × I.
[d] Tested against H × S × I.

TABLE 3

ANALYSIS OF VARIANCE OF KINESTHETIC ADJUSTMENT OF ROD TO VERTICALITY UNDER BODY TILT (EXP. II)

| Source of Variance | *df* | Sum of Squares | *F* | *p* |
|---|---|---|---|---|
| Body position (B) | 2 | 1786.57 | 19.91[a] | <.01 |
| Starting position (S) | 2 | 21.46 | 1.24[b] | >.05 |
| Individuals (I) | 20 | 369.56 | <1.00[a] | >.05 |
| B × S | 4 | 34.64 | 2.81[c] | <.05 |
| B × I | 40 | 1794.54 | 14.56[c] | <.01 |
| S × I | 40 | 153.65 | 1.25[c] | >.05 |
| B × S × I | 80 | 246.25 | | |
| Total | 188 | 4406.67 | | |

[a] Tested against B × I.
[b] Tested against B × S.
[c] Tested against B × S × I.

the body is tilted to the right, the mean position of the apparent vertical is rotated to the left (−2.9°) of control (1.6°), and when the body is tilted to the left, the mean position of the apparent vertical is to the right (4.6°) of control. The positions of the apparent vertical under all three test conditions differ significantly from each other ($p < .01$). The analysis of variance is summarized in Table 3.

*Experiment III: Body rotation.*—When *S* is rotated around his own vertical axis there are significant effects of labyrinthian stimulation on the kinesthetic perception of verticality. These effects are summarized in Table 4. Under control conditions (no rotation) the apparent vertical is closest to objective vertical (−.5°). Under CW acceleration (A–CW) the apparent vertical is tilted to the left (−2.6°) of control; under CCW acceleration (A–CCW) the apparent vertical is tilted to the right (.8°) of control; under deceleration from CW rotation (D–CW) the apparent vertical is tilted to the right (1.3°) of control; and under deceleration from CCW rotation (D–CCW) the apparent vertical is tilted to left (−3.2°) of control. Thus the apparent vertical is shifted opposite the direction of acceleration, i.e., when acceleration is to the right, the tilt of the rod in the fronto-parallel plane is to the left, and vice versa.[2] Moreover

[2] As mentioned in (2), this statement is dynamically but not geometrically correct. That is, CW and CCW with reference to rotation deal

TABLE 4

KINESTHETIC ADJUSTMENT OF ROD TO VERTICALITY UNDER BODY ROTATION (EXP. III)

*F tests*

| Source of Variance | *df* | Sum of Squares | *F* |
|---|---|---|---|
| Conditions (C) | 4 | 95.34 | 20.03*** |
| Individuals (I) | 5 | 13.97 | 2.34 |
| Error (C × I) | 20 | 23.80 | |
| Total | 29 | 133.11 | |

*t tests* (Mean differences)

| Conditions | Mean | A–CW | D–CW | A–CCW | D–CCW |
|---|---|---|---|---|---|
| Control (C) | −.5 | 2.1** | 1.8** | 1.3* | 2.6*** |
| Acceleration CW (A–CW) | −2.6 | | 3.9*** | 3.4*** | .5 |
| Deceleration from CW rotation (D–CW) | +1.3 | | | .5 | 4.4*** |
| Acceleration CCW (A–CCW) | +.8 | | | | 3.9*** |
| Deceleration from CCW rotation (D–CCW) | −3.1 | | | | |

* Approaches significance ($p = .06$).
** Significant at or below the .05 level of confidence.
*** Significant at or below the .01 level of confidence.

deceleration from rotation in one direction functions like acceleration in the opposite direction.

The *t* tests in Table 4 show that the mean position of apparent vertical without stimulation (control) is significantly different from the apparent vertical under each of the experimental test conditions.[3] Acceleration in one direction differs significantly from acceleration in the other direction; this holds also for deceleration. Acceleration in one direction differs significantly from deceleration in the opposite direction. However, in both cases acceleration in one direction does not differ significantly from deceleration from rotation in the other direction.

## Discussion

The main aim of this study was to inquire into the generality of the effects of extraneous stimulation on object perception which have been previously demonstrated on the visually perceived vertical (**1, 2, 4, 6**). In the previous studies it has been found that there is a specific effect of extraneous stimulation when asymmetrically applied. If an extraneous stimulus (e.g., auditory, labyrinthian, etc.) is applied to one side of the body, a visual line appears vertical when rotated away from the side stimulated. Related results were obtained with stimulation ensuing from body tilt: when the body is tilted slightly to one side, the position of the apparent vertical (visual) is tilted to the other side.

Verticality is a spatial property of an object, which can be experienced by kinesthetic as well as by visual cues. One may, therefore, utilize this property to find out whether the effects of extraneous stimulation and body tilt are specific to visual object perception or whether they are of a more general nature. The significance of the present experiments is the demonstration that for kinesthetic perception of the vertical results have been obtained essentially identical in nature with those previously reported for visual perception of the vertical. That is, when *S* is blindfolded, the kinesthetically perceived apparent vertical is in a position tilted to the left when the head is tilted to the right, when the body is tilted to the right, and when the body is accelerated to the right or decelerated to the left. Analogous relations hold for the opposite test conditions.

These results suggest that interpretations which do not depend on the specific characteristics of one or the other modality are more satisfactory than those which take into account only a specific modality—visual or kinesthetic but not both.

Such specific interpretations of perceptual change are, for instance, those which deal with retinal displacements subsequent to compensatory torsion of the eye, compensatory eye movements, nystagmus, etc. We do not deny that ocular-muscular changes involved in eye-movements, nystagmus, etc., are of significance. On the contrary, by their very nature as part of the organismic state, they must be considered. However, there is a difference whether the interpretation is formulated in terms of eye movements as ocular-muscular events or in terms of eye movements as conditions for retinal displacements.

In contradistinction to specific visual theories our own interpretation, because of its organismic nature, is general enough to take care of results

with motion around the vertical axis of the body, whereas left and right deals with rotation in the fronto-parallel plane.

[3] The *p* value (.06) for the difference between A–CCW and control approaches significance.

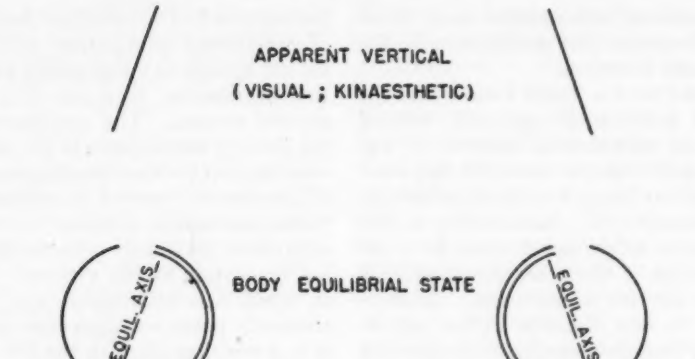


FIG. 1. Hypothetical schema: Relation between position of apparent vertical and body equilibrium under various types of extraneous stimulation

irrespective of the sense modality involved in the perception.

In particular, the sensory-tonic field theory of perception fulfills the generality requirement because the same hypothetical mechanism is applicable for experimental results on kinesthetic as well as on visual perception of verticality. This mechanism is schematically presented in Fig. 1. There are two assumptions: (*a*) A line is perceived as vertical when its position in space is in tune with the existing body-equilibrial state. (*b*) Asymmetrically applied extraneous stimulation including body tilt shifts the body-equilibrial axis away from the side of stimulation. With extraneous stimulation the body-equilibrial state, represented by the equilibrial axis, is that depicted in Fig. 1A and 1B; there the position of the apparent vertical must be displaced in the direction opposite the side of stimulation to become commensurate with the changed body-equilibrial axis.[4]

Finally, it should be emphasized that the present experiment provides new evidence for the validity of the construct of "functional equivalence." Labyrinthian stimulation and body tilt are factors affecting verticality in an equivalent way, independent of whether the affected object is perceived kinesthetically or visually.

## SUMMARY

The present experiment was carried out to study the effect of head tilt, body tilt, and concentric body rotation on the kinesthetic perception of

[4] The mechanism hypothesized here has been discussed in (4) and will be discussed in greater detail in a forthcoming theoretical paper (5).

verticality. The *S*, while blindfolded, had the task of adjusting a rod pivoted at the lower end, to the vertical by moving the fingers of both hands along the rod.

The results are essentially identical with those previously reported for visual perception of the vertical. That is, the position of the rod that appeared vertical was objectively tilted to the left when: (*a*) the head was tilted right; (*b*) the body was tilted right; (*c*) the body was accelerated around the vertical axis in the clockwise direction; (*d*) the body was decelerated from constant counterclockwise rotation. For the opposite test conditions the respective displacements were in a direction opposite to those mentioned above.

The significance of this study consists in the demonstration that the effects of extraneous stimulation and body tilt previously described for visual perception of verticality can be generalized to include kinesthetic perception of verticality. The results, thus, broaden the interpretations given within the sensory-tonic field theory of perception.

(Received for priority publication April 10, 1952)

## References

1. Wapner, S., Werner, H., & Chandler, K. A. Experiments on sensory-tonic field theory of perception: I. Effect of extraneous stimulation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 341–345.
2. Wapner, S., Werner, H., & Morant, R. B. Experiments on sensory-tonic field theory of perception: III. Effect of body rotation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 351–357.
3. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.*, 1949, **18**, 88–107.
4. Werner, H., & Wapner, S. Experiments on sensory-tonic field theory of perception: IV. Effect of initial position of a rod on apparent verticality. *J. exp. Psychol.*, 1952, **43**, 68–74.
5. Werner, H., & Wapner, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–338.
6. Werner, H., Wapner, S., & Chandler, K. A. Experiments on sensory-tonic field theory of perception: II. Effect of supported and unsupported tilt of the body on visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 346–350.

# THRESHOLDS OF THE HUMAN EYE FOR ELECTRIC STIMULATION BY DIFFERENT WAVE FORMS[1]

J. W. GEBHARD

*Applied Physics Laboratory, The Johns Hopkins University*

Much progress in visual science has been made by measuring the electric response of the eye to a light stimulus (**9, 10**). Less deeply explored has been the possibility of revealing some of the mechanisms of vision by the reverse process. This is to use an electric input and to measure the response in terms of the sensation of light such a stimulus calls forth. Investigators have shown sporadic interest in this matter ever since LeRoy (**15**) discovered in 1755 that an electric pulse to the eye will easily produce a bluish-white sensation of light. Visual perceptions caused by inadequate stimuli are called *phosphenes* and the electrical phosphenes have been studied in a variety of respects for many years. Most of the early work was qualitative, but since 1900 measurements have frequently appeared. In general, there has been poor agreement among the findings reported and the history of inconsistency continues into the research of the present.

A new and perhaps very important turn has been given this work by the studies of Motokawa and his associates (**1, 8, 16, 17, 25**). These workers have been ingenious in their use of the threshold of an electrical phosphene as a means of analyzing the effect of photic stimulation on the activity of the visual organ. In a formidable output of some 40 papers in the last four years, they have touched upon the electrophysiology of brightness and color discrimination, color deficiency, contrast, spatial summation, inhibition, visual illusions, and flicker effects (**8**). This work is of considerable interest and is a fresh approach to the physiological basis of vision. Nevertheless, in spite of the promising nature of the published results, it is much too early to characterize them firmly. One can only say now that a great deal of verification must be done before Motokawa's findings are wholly accepted.

The main concern of this paper is with a problem in electrostimulation in which certain results of Motokawa disagree with those of other workers. The problem is: Does the stimulus strength-frequency relation have one or several minima? This relationship is found by measuring the electrical thresholds of just noticeable flicker phosphenes for different frequencies of intermittent electric input. It has long been known from work on other tissues that there are optimum frequencies of excitation at which thresholds are lowest (**13, 14**). When threshold current $I$ is plotted against log frequency $n$, U-shaped curves result that are symmetrical about the optimum frequency (**5**). The theoretical importance of the $I - \log n$ curves has been discussed by Hill (**11**), Coppée (**6**), Katz (**12**), Schaefer (**19**), and others.

For the human eye, Rohracher (**18**) showed in 1935 that the strength-frequency curve as determined by sine waves passed through a distinct mini-

[1] This research was done while the author was on the staff of The Johns Hopkins University Institute for Cooperative Research, Baltimore, Md. It was carried out under Contract N5-ori-166, Task Order I, between the Special Devices Center, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-145, Project Designation No. NR-784-001 under that contract.

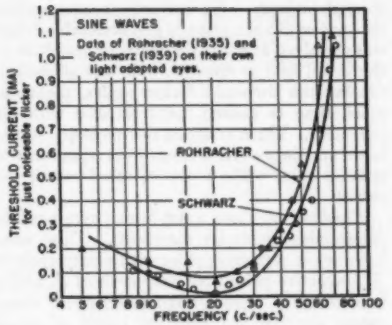
SINE WAVES
Data of Rohracher (1935) and Schwarz (1939) on their own light adapted eyes.
ROHRACHER
SCHWARZ
THRESHOLD CURRENT (MA) for just noticeable flicker
FREQUENCY (c./sec.)

FIG. 1. Strength-frequency ($I - \log n$) curves showing a single optimum frequency (after Rohracher and Schwarz)

mum at about 20 cy./sec. for the light-adapted condition. Schwarz (**20, 21, 22, 23, 24**), in a series of careful and varied experiments between 1936 and 1947, verified and extended Rohracher's finding of only one frequency of maximum stimulus effectiveness for the light-adapted eye. Figure 1 shows the data of these investigators.

On the other side of the argument, Barnett and Posner (**3**), Motokawa and Iwama (**17**), and Abe (**1**) found the relationship to have several minima for the light-adapted eye. Barnett and Posner used sine waves and reported a minimum at 22 cy./sec. for a "bright" condition of light adaptation and at 7 cy./sec. for a "dim" condition. Motokawa and Iwama, who used both sine waves and rectangular pulses, reported minima at 18, 37, and 53 cy./sec.[2] Abe also worked with rectangular pulses and claimed minima at 7, 20, 35, 54, and 74 cy./sec. Figure 2 shows results of this kind. The data of Fig. 1 and 2 are clearly not in agreement.

A second concern of this paper is

[2] In another place Motokawa (**16**) also reported data on two minima while mentioning three and four.

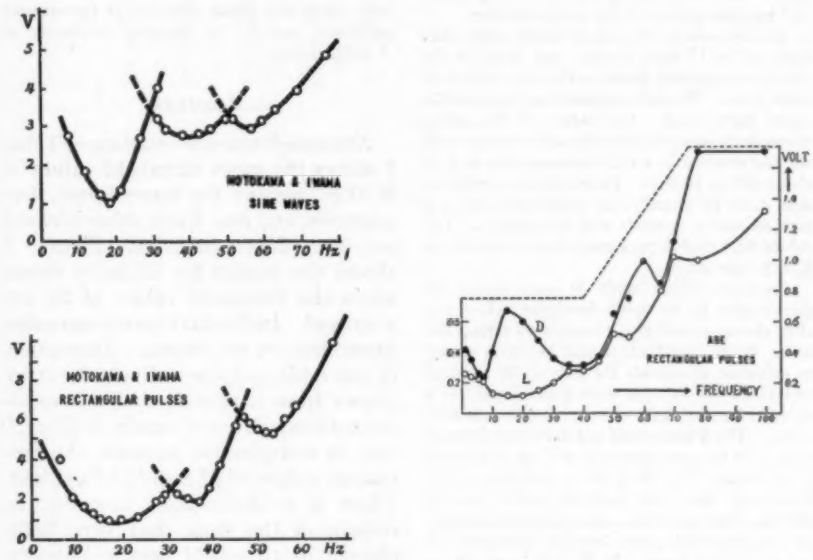


FIG. 2. Strength-frequency curves showing several optimum frequencies. Curves L and D are for light- and dark-adapted conditions, respectively. (By permission of the *Tohoku J. exp. Med.*)

with how the stimulus effectiveness of an intermittent electric input to the eye is affected by polarity and wave form. The problems of stimulus effectiveness and the strength-frequency relationship were studied at the same time in experiments conducted in 1948 before seeing the papers from Motokawa's laboratory (7). This is unfortunate with respect to the first problem in that insufficient frequencies were chosen above 25 cy./sec. to evaluate adequately the question of minima in this region.

## Method

*Stimulators.*—Two electronic stimulators were used: a conventional sine wave oscillator and a rectangular pulse generator specially constructed for safe use on the eye (2). The second generator provided anodal and cathodal pulses, or a train of pulses alternating in polarity. These were of controllable repetition rate, duration, and amplitude. Anodal or cathodal square waves were produced by adjusting the required polarity to a duty cycle (D/T) of .5. A "symmetrical" square wave was formed by alternating positive and negative pulses of the same duration.

*Measurements.*—A nonpolarizable silver electrode, 17 × 17 mm. in size, was fixed to the glabella equidistant between the eyes with electrode paste. The other contact was made to the entire right hand. Application of the active electrode by scraping the skin and cleaning with acetone resulted in a DC resistance due to *S* of about 800 to 1000 Ω. Threshold measurements were made by means of an oscillograph across a high resistance in series with the subject. The values were read in peak current and equated for R.M.S. current.

*Procedure.*—Four adult *S*s were used. An examination by an ophthalmologist before and after electro-stimulation revealed no retinal defects. Since visual adaptation has been shown to influence thresholds for electrically induced sensations, the *S*s' eyes were kept adapted to a constant level of light while measurements were taken. The *S* was seated in a dark experimental room with his eyes open and as close as possible to the center of a 36 × 36-in. diffusing screen illuminated from the back by white light of 3600° K. This provided a reasonably homogeneous adapting field at two levels of luminance: .1 and 10 millilamberts. At *S*'s left hand was a knob that he used to control the current passing through his body. His instructions were to adjust the knob until a condition of just noticeable flicker was perceived at either the center or periphery of the visual field as directed. He was allowed to bracket the threshold flicker region, taking as much time as desired, until satisfied with his judgment. The *E* arranged the frequency, waveform, and duty-cycle settings from equipment in another room while *S* was switched out of the circuit.

*Conditions.*—Four wave forms were used in determining the flicker thresholds. These were a sine wave, a positive pulse with a duty cycle of .5, a negative pulse with duty cycle of .5, and a "symmetrical" square wave made up of alternating positive and negative pulses of equal duration. Thresholds were obtained at 12 frequencies for each wave form except the sine wave. These were at 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 35, 45, and 60 cy./sec. For the sine wave, the frequencies at 5 and 7.5 cy./sec. were omitted because of limitations of the oscillator. The frequencies were presented in a random order until an equal number of measurements had been obtained at each frequency. The data are means based on either five or ten thresholds as indicated in the next section. Most of the thresholds were obtained for flicker observed in the central 5° of the visual field at the adapting luminance of 10 millilamberts. For purposes of comparison, however, some data were taken with flicker observed at the extreme periphery, and for an adapting luminance of .1 millilambert.

## Results

*Strength-frequency relation.*—Table 1 shows the mean threshold values in R.M.S. current for wave forms, frequencies, and *S*s. Each value is based on five determinations. Figure 3 shows the results for all wave forms when the threshold values of *S*s are averaged. Individual curves were also drawn but are not shown. Inspection of the table and the individual curves drawn from it reveals that one minimum threshold, very nearly at 20 cy./sec., is conspicuous between the frequency values of 17.5 and 22.5 cy./sec. There is evidence also, however, in certain of the data that very little change in threshold occurs between some of the individual points meas-

ured between 7.5 and 22.5 cy./sec. There are several instances of no change and two cases, both for Subject C, of reversals. The extent to which these variations are "minima" is questionable and is discussed below.

*Stimulus effectiveness.*—Data on the problem of wave-form effectiveness

TABLE 1

EFFECT OF POLARITY AND WAVE FORM ON THRESHOLDS OF JUST NOTICEABLE FLICKER OBSERVED IN THE CENTER OF THE FIELD

Values are in mA equated for R.M.S. current. The luminance of the adapting field was 10 mL.

Symmetrical Square Wave

| Subject | Frequency in cy./sec. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 7.5 | 10 | 12.5 | 15 | 17.5 | 20 | 22.5 | 25 | 35 | 45 | 60 |
| A | .84 | .76 | .66 | .50 | .34 | .24 | .28 | .32 | .36 | .40 | .82 | 1.74 |
| B | .36 | .26 | .24 | .16 | .16 | .08 | .08 | .12 | .14 | .20 | .26 | .68 |
| C | .90 | .48 | .38 | .38 | .40 | .30 | .30 | .36 | .40 | .60 | .90 | 1.48 |
| D | .78 | .62 | .54 | .48 | .38 | .34 | .34 | .36 | .40 | .54 | .66 | 1.08 |
| Average | .72 | .52 | .46 | .38 | .32 | .24 | .24 | .28 | .32 | .44 | .66 | 1.24 |

Anodal Square Wave

| Subject | Frequency in cy./sec. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 7.5 | 10 | 12.5 | 15 | 17.5 | 20 | 22.5 | 25 | 35 | 45 | 60 |
| A | .73 | .71 | .68 | .49 | .37 | .27 | .33 | .39 | .38 | .42 | .83 | 1.01 |
| B | .29 | .19 | .13 | .10 | .09 | .05 | .05 | .06 | .08 | .12 | .18 | .34 |
| C | .78 | .42 | .40 | .41 | .39 | .31 | .28 | .34 | .35 | .57 | .85 | 1.04 |
| D | .56 | .45 | .41 | .36 | .31 | .33 | .31 | .39 | .40 | .46 | .57 | .78 |
| Average | .59 | .44 | .40 | .34 | .29 | .24 | .24 | .30 | .30 | .39 | .61 | .79 |

Cathodal Square Wave

| Subject | Frequency in cy./sec. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 7.5 | 10 | 12.5 | 15 | 17.5 | 20 | 22.5 | 25 | 35 | 45 | 60 |
| A | .80 | .59 | .49 | .40 | .30 | .22 | .20 | .26 | .30 | .34 | .64 | .95 |
| B | .21 | .14 | .13 | .10 | .07 | .07 | .06 | .06 | .08 | .13 | .18 | .45 |
| C | .75 | .45 | .42 | .41 | .41 | .32 | .31 | .31 | .36 | .62 | .78 | .98 |
| D | .38 | .35 | .26 | .22 | .18 | .18 | .19 | .22 | .26 | .26 | .33 | .61 |
| Average | .54 | .38 | .32 | .28 | .24 | .20 | .19 | .21 | .25 | .34 | .48 | .75 |

Sine Wave

| Subject | Frequency in cy./sec. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 7.5 | 10 | 12.5 | 15 | 17.5 | 20 | 22.5 | 25 | 35 | 45 | 60 |
| A | | | .80 | .59 | .43 | .25 | .21 | .25 | .27 | .34 | .61 | 1.14 |
| B | | | .18 | .13 | .11 | .10 | .06 | .09 | .11 | .14 | .22 | .67 |
| C | | | .33 | .28 | .32 | .28 | .20 | .25 | .27 | .34 | .66 | .95 |
| D | | | .51 | .43 | .39 | .33 | .27 | .37 | .37 | .42 | .56 | .73 |
| Average | | | .45 | .36 | .31 | .24 | .18 | .24 | .25 | .31 | .52 | .88 |

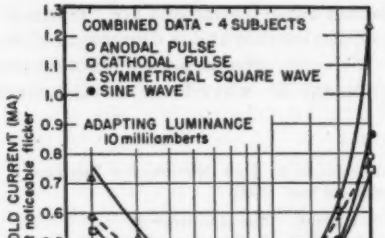


FIG. 3. $I - \log n$ curves for different wave forms (from Table 1)

makes it apparent that there is no difference between one polarity or wave form and another when the measures are equated for R.M.S. current. Square waves yield lower thresholds than sine waves by an amount predictable from the total AC power. The result that AC power controlled stimulus effectiveness was checked by separate measurements on the effect of duty cycle. The upper section of Table 2 shows complete measurements taken on the anodal pulse at 30 cy./sec. Each value is the mean of ten measurements. Figure 4 gives the data in graphic form. Both anodal and cathodal pulses were further measured with a frequency of 35 cy./sec. at duty cycles of .3, .5, and .7. These results are also based on ten determinations and are given in the rest of the table. The data clearly show that pulses are most effective when the duty cycle is .5; i.e., when the pulse becomes a square wave and the total AC power is maximum. These findings indicate that the effec-

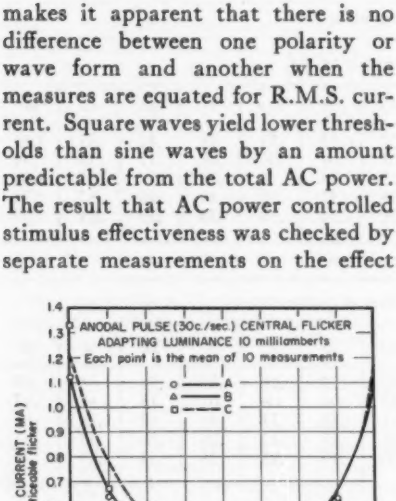


FIG. 4. Stimulus effectiveness related to the duty cycle of rectangular pulses (from Table 2)

TABLE 2

EFFECT OF DUTY CYCLE ON THE STIMULUS EFFECTIVENESS OF ANODAL AND CATHODAL RECTANGULAR PULSES

Thresholds for just noticeable flicker in mA. The luminance of the adapting field was 10 mL.

Anodal Pulse at 30 cy./sec.

| Subject | Duty Cycle | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | .1 | .2 | .3 | .4 | .5 | .6 | .7 | .8 |
| A | 1.12 | .64 | .48 | .43 | .41 | .43 | .51 | .63 |
| B | .32 | .20 | .16 | .12 | .10 | .11 | .13 | .16 |
| C | 1.33 | .67 | .60 | .46 | .43 | .46 | .48 | .62 |

Anodal Pulse at 35 cy./sec.

| Subject | Duty Cycle | | |
|---|---|---|---|
| | .3 | .5 | .7 |
| A | .47 | .35 | .40 |
| B | .12 | .10 | .13 |
| C | .50 | .41 | .45 |

Cathodal Pulse at 35 cy./sec.

| Subject | Duty Cycle | | |
|---|---|---|---|
| | .3 | .5 | .7 |
| A | .56 | .43 | .52 |
| B | .11 | .08 | .10 |
| C | .58 | .52 | .68 |

tive stimulus for the eye of an intermittent electric input is alternating current and that the stimulus effectiveness of any wave form is dependent upon the total power of the AC component.

## Discussion

The general result of these measurements is that the visual mechanism shows only one frequency of electric input at which the threshold for flicker is conspicuously a minimum. This finding is in satisfactory agreement with the results of Rohracher and of Schwarz. It is also in agreement with certain theoretical considerations of nerve response to alternating currents proposed by Hill, Coppée, and Schaefer that predict symmetrical, U-shaped $I - log\ n$ curves. It appears not to agree with Barnett and Posner, Motokawa and Iwama, or Abe, all of whom have found several well-defined minima in the strength-frequency curve. There is, nevertheless, a suggestion that the $I - log\ n$ curves obtained in this study are not as symmetrical as might be desired. However, the irregularities in the curves, as measured in milliamperes, are very small. They reflect, if anything, the small number of measurements taken at each frequency rather than the presence of true minima.

Barnett and Posner did not report the extent of their data and merely indicated that three *Ss* were used for the "bright" adaptation level and two for the "dim." Neither did they state where, in the visual field, flicker was judged. This, taken with a rather crude means of controlling adapting luminance, makes their published curves difficult to evaluate. Their "bright" minimum at 22 cy./sec. corresponds fairly well with data from Rohracher and Schwarz. The "dim" minimum at 7 cy./sec. agrees only with Abe's finding.

Motokawa and Iwama measured voltage rather than current. This is perhaps open to the objection that variations of resistance due to changes in the tissues, sweating, loosening of the electrodes and the like, may unduly affect voltage readings at threshold. In addition, their means of controlling adaptation was similar to that of Barnett and Posner. No indication is given of the amount of data used in constructing their curves and it may be noted that their results on sine waves (Fig. 2) show two quite different values of threshold plotted at the frequency of 30 cy./sec. In this figure they have carried the limb of one curve through one value and that of the adjacent curve through the other value. Unless this is an error, there is some question of the justification for this. Finally, the authors objected that Rohracher and Schwarz missed two additional minima because of an insufficient number of frequencies measured. But Schwarz (**22, 23**) measured the same number of points between 10 and 55 cy./sec. as did Motokawa and Iwama. They were not the same points, although they were equally well dispersed along the frequency range.

Abe's method of achieving a light-adapted eye was to pre-illuminate with 24,000 lux (meter candles) for an unspecified length of time and to measure thresholds after 1.5 min. had elapsed. This means that he was working in an early stage of dark adaptation rather than in a condition of maintained light adaptation. Again, voltage was measured using about the same technique as that of Motokawa and Iwama. It is interesting to note, however, that Abe's thresholds were only about one-tenth as large. No details were given as to the amount of data obtained. More, it is difficult to see in Abe's graphs for the light-adapted eye (L) a clear indication of the five minima he reports. Abe's points appear to be fitted reasonably well by a smooth curve with a single minimum at 20 cy./sec. His curve for the dark-adapted eye (D) which shows a *high* threshold at 20 cy./sec. is completely at variance with evidence from Schwarz (**24**). Abe attributed the absence of the 20 cy./sec. minimum to the low electrical excitability of the rods in the dark-adapted condition. This hypothesis was studied in some detail by Tukahara and Abe (**25**) who presented data to show that rod sensitivity to electric stimulation at 20 cy./sec. disappeared when light was removed from the retina.

As pointed out earlier, the frequencies above 25 cy./sec. measured in the present study were too far apart to disclaim the presence of additional minima beyond this frequency. It can only be noted that the threshold values at 25, 35, 45, and 60 cy./sec. are reasonably well fitted by a smooth curve. Schwarz, at different times,

has measured 14 to 18 frequencies distributed between 25 and 110 cy./sec. and his values are likewise satisfactorily fitted by a smooth curve.

There are many reasons why marked variation can result in the measures of thresholds such as these. Among the obvious difficulties are those related to intra-*S* variability and the technical problems of consistent electrostimulation and measurement. In addition, there are two other sources of difficulty. The first results from the fact that the electrical thresholds depend upon the adaptation level of the eyes. The effect of adapting luminance is seen in Fig. 5 where threshold measures obtained at 10 and .1 millilamberts are compared. It is apparent that any inadvertent change in adaptation level will produce a shift in the thresholds over much of the frequency range. This same effect has been shown in the 1947 data of Schwarz (**24**).

The second source of potential trouble lies in controlling where, in the visual field, *S* looks for the flicker. If his judgments are made in the central 5° of the field, the thresholds will be higher than if judgments are made elsewhere. A comparison of the center and the extreme periphery is made in Fig. 6. The reason for this difference is not clear. It cannot be concluded from these measures that the thresholds at the center and the periphery differ because of regional differences in the visual mechanism. The method of controlling adapting luminance used in this study resulted in illuminating the periphery of the retina more weakly than the center. This would tend to reduce the thresholds at the periphery, assuming an equal electrical excitability of all areas of the eyes. However, Bourguignon and Déjean (**4**), using the entirely different method of stimulating the eye with condenser discharges and measuring chronaxies, found that peripheral phosphenes were easier to arouse

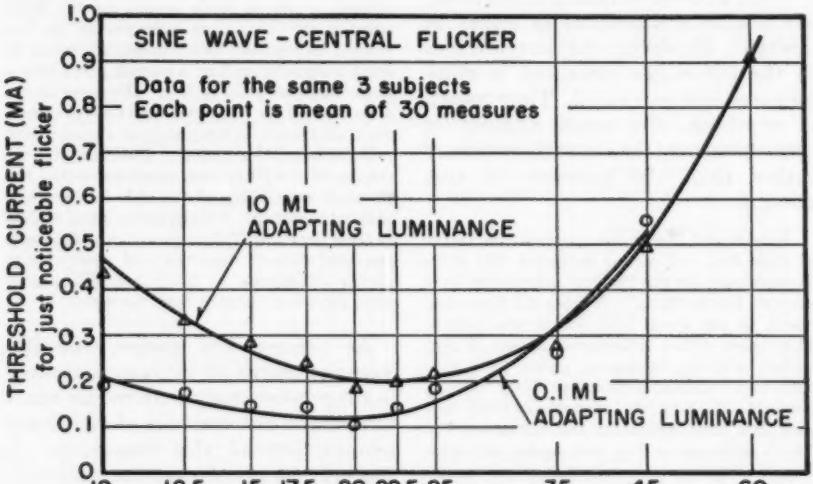


FIG. 5. Effect of adapting luminance on the $I - \log n$ curve

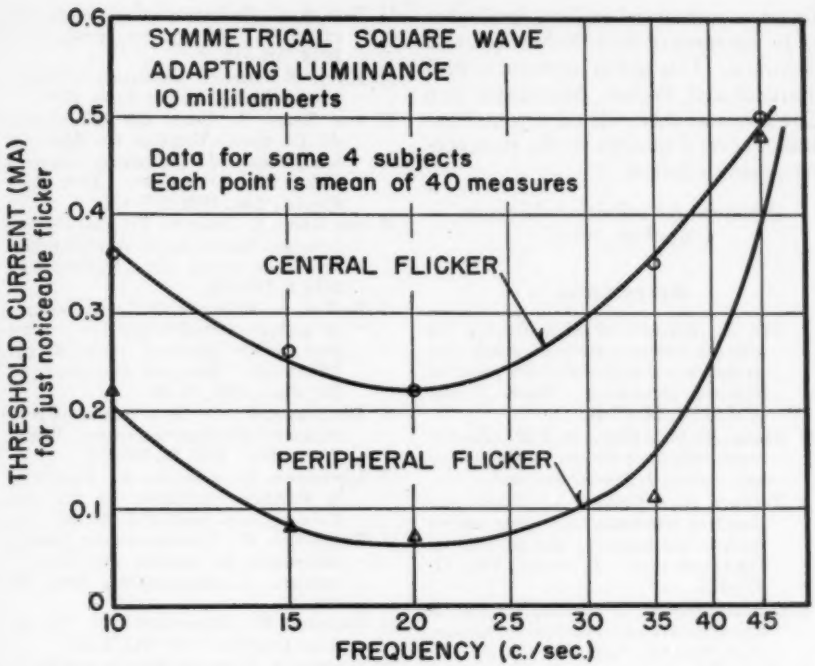


FIG. 6. Effect of position in visual field where flicker is judged on the $I - \log n$ curve

than central ones. They reported chronaxies of 1–2 msec. for peripherally localized sensations and 2–3 msec. for those in the center. The significant fact for this discussion is that if *S* judged the flicker first at one place, then at another in the visual field, varying values of stimulus intensity would be found. Inadequate control of both adaptation level and of where flicker is judged, therefore, could conceivably produce "minima" where none exist.

## SUMMARY

Threshold intensities of the intermittent electric excitation required for *S*s to experience just noticeable flicker were obtained for light-adapted eyes with sine waves, anodal and cathodal square waves and rectangular pulses.

1. Polarity was found to be unimportant. Anodal or cathodal stimuli were equally effective in producing an electrically aroused phosphene.

2. Square waves gave lower thresholds than sine waves by an amount predictable from the total AC power.

3. Rectangular pulses were most effective when the duty cycle was .5; i.e., when the pulse became a square wave and the total AC power was maximum. These results indicate that the effective intermittent electric stimulus for the eye is alternating current and that the efficiency of any wave form is dependent upon the total power of the AC component.

4. Thresholds for all wave forms were lowest at a frequency of about 20 cy./sec. This frequency was the only one at which an unequivocal min-

imum was observed and is a result that is in agreement with Rohracher and Schwarz. It is not in agreement with Barnett and Posner, Motokawa and Iwama, and Abe, all of whom have found several minima in the strength-frequency relation.

(Received for priority publication April 22, 1952)

## References


1. Abe, Z. Influence of adaptation on the strength-frequency curve of human eyes as determined with electrically produced flickering phosphenes. *Tohoku J. exp. Med.*, 1951, **54**, 37–44.
2. Barger, D. M., & Gebhard, J. W. An electronic stimulator for use on the human eye. *Amer. J. Psychol.* (In press.)
3. Barnett, A., & Posner, M. Effects produced by impressing alternating electric fields of low frequency and intensity on the human skull. *J. Psychol.*, 1941, **11**, 271–278.
4. Bourguignon, G., & Déjean, R. Double chronaxie du système optique de l'homme. *C. R. Acad. Sci., Paris*, 1925, **180**, 169–172.
5. Coppée, G. La pararésonance dans l'excitation par les courants alternatifs sinusoïdaux. *Arch. int. Physiol.*, 1934, **40**, 1–58.
6. Coppée, G. Stimulation by alternating currents. *Cold Spr. Harb. Sympos. quant. Biol.*, 1936, **4**, 150–162.
7. Gebhard, J. W. Electrical stimulation of the eye by sine waves, square waves and rectangular pulses. *Amer. Psycholologist*, 1949, **4**, 358 (Abstract).
8. Gebhard, J. W. Review of Motokawa's work on electrical excitation of the human eye. *Minutes 29th Meeting Armed Forces-NRC Vision Committee* 16–17 November 1951, National Research Council, Washington, D. C.
9. Granit, R. *Sensory mechanisms of the retina.* London: Oxford Univer. Press, 1947.
10. Hartline, H. K. Nerve messages in the fibres of the visual pathway. *J. opt. Soc. Amer.*, 1940, **30**, 239–247.
11. Hill, A. V. Excitation and accommodation in nerve. *Proc. roy. Soc. Lond.*, 1936, **B119**, 305–355.
12. Katz, B. *Electric excitation of nerve.* London: Oxford Univer. Press, 1939.
13. von Kries, J. Ueber die Abhängigkeit der Errengungs-Vorgänge von dem zeitlichen Verlaufe der zur Reizung dienenden Elektricitäts-Bewegungen. *Arch. Anat. Physiol., Lpz.*, 1884, 337–372.
14. von Kries, J. Ueber die Erregung des motorischen Nerven durch Wechselströme. *Ber. Verh. naturf. Ges., Freiburg i.B.*, 1884, **8**, 170–205.
15. Le Roy, ——(Mémoire) Où l'on rend compte de quelques tentatives que l'on a faites pour guérir plusieurs maladies par l'Électricité. *Mém. mat. phys. Acad. roy. Sci., Paris*, 1755, 60–98.
16. Motokawa, K. On the mechanism of periodic excitability of nervous tissue. *Tohoku J. exp. Med.*, 1949, **50**, 307–318.
17. Motokawa, K., & Iwama, K. Resonance in electrical stimulation of the eye. *Tohoku J. exp. Med.*, 1950, **53**, 201–206.
18. Rohracher, H. Ueber subjective Lichterscheinungen bei Reizung mit Wechselströmen. *Z. Sinnesphysiol.*, 1935, **66**, 164–181.
19. Schaefer, H. *Electrophysiologie.* Vienna: Franz Deuticke, 1940–1942, 2 vols.
20. Schwarz, F. Ueber die Wirkung von Wechselstrom auf das Sehorgan. *Z. Sinnesphysiol.*, 1936–38, **67**, 227–244.
21. Schwarz, F. Ueber die Reizung des Sehorgans durch niederfrequente elektrische Schwingungen. *Z. Sinnesphysiol.*, 1939–40, **68**, 92–118.
22. Schwarz, F. Quantitive Untersuchungen über die optische Wirkung sinusförmiger Wechselströme. *Z. Sinnesphysiol.*, 1940, **69**, 1–26.
23. Schwarz, F. Ueber die Reizung des Sehorgans durch doppelphasige und gleichgerichtete elektrische Schwingungen. *Z. Sinnesphysiol.*, 1941, **69**, 158–172.
24. Schwarz, F. Ueber die elektrische Reizbarkeit des Auges bei Hell- und Dunkeladaptation. *Pflüg. arch. ges. Physiol.*, 1947, **249**, 76–86.
25. Tukahara, S., & Abe, Z. Resonance phenomena of photopic and scotopic receptors. *Tohoku J. exp. Med.*, 1951, **54**, 189–196.

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

## The issues listed below are available for sale

| Year | Volume | Available Numbers | | | | | | Price per Number | Price per Volume |
|---|---|---|---|---|---|---|---|---|---|
| 1916 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1917 | 2 | 1 | 2 | – | 4 | 5 | – | $1.25 | $5.00 |
| 1920 | 3 | – | 2 | – | 4 | 5 | – | $1.25 | $3.75 |
| 1921 | 4 | 1 | – | 3 | 4 | 5 | 6 | $1.25 | $6.25 |
| 1922 | 5 | – | – | 3 | 4 | 5 | – | $1.25 | $3.75 |
| 1923 | 6 | – | – | – | – | – | – | —— | —— |
| 1924 | 7 | 1 | 2 | – | – | 5 | – | $1.25 | $3.75 |
| 1925 | 8 | 1 | 2 | 3 | 4 | – | 6 | $1.25 | $6.25 |
| 1926 | 9 | – | 2 | 3 | – | 5 | 6 | $1.25 | $5.00 |
| 1927 | 10 | – | – | 3 | 4 | 5 | 6 | $1.25 | $5.00 |
| 1928 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1929 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1930 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1931 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1932 | 15 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1933 | 16 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1934 | 17 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1935 | 18 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1936 | 19 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1937 | 20 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 21 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1938 | 22 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 23 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1939 | 24 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 25 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1940 | 26 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 27 | 1 | 2 | 3 | 4 | 5 | – | $1.25 | $6.25 |
| 1941 | 28 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 29 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1942 | 30 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 31 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1943 | 32 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 33 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1944 | 34 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1945 | 35 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1946 | 36 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1947 | 37 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1948 | 38 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1949 | 39 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1950 | 40 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1951 | 41 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 42 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1952 | 43 | By subscription, $14.00 (2 vols.) | | | | | | $1.25 | —— |
| | 44 | | | | | | | | |
| | | List price, Volumes 1 through 42 | | | | | | | $269.00 |
| | | 30% discount | | | | | | | 80.70 |
| | | Net price, Volumes 1 through 42 | | | | | | | $188.30 |

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

*Address orders and subscriptions to*

## AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.